UNIVERSITATEA DE VEST DIN TIMIŞOARA
FACULTATEA DE SOCIOLOGIE ŞI PSIHOLOGIE

Lector Mihai HOHN Prep. Laura MĂRUŞTER

Curs de statistică aplicată în ştiinţe sociale

Vol. I

ANDRIEŞ C. LUCIAN
Nr 2774
Biblioteca Personală

- pentru uzul studenţilor -

Timişoara, 1998

Cuprins

INTRODUCERE 3

CAPITOLUL 1 NOŢIUNI GENERALE 5

CAPITOLUL 2 SOFTWARE STATISTIC 8

2.1. SPSS - STATISTICAL PACKAGE FOR SOCIAL SCIENCE 11
2.1.1 Primii paşi *12*
2.1.2.Proiectarea unui studiu *12*
2.1.3. Date în SPSS *15*

CAPITOLUL 3 STATISTICI DESCRIPTIVE 20

3.1. GRUPAREA DATELOR 22
3.2. REPREZENTĂRI GRAFICE 25
3.2.1 Diagramele *25*
3.2.2. Histograma *28*
3.2.3. Poligonul frecvenţelor *30*
3.2.4. Curba frecvenţelor cumulate *32*
3.2.5. Curba procentajelor cumulate *34*
3.3. PARTICULARITĂŢI ALE DISTRIBUŢIILOR STATISTICE 36
3.4. CONSIDERAŢII PRIVIND CONSTRUCŢIA ŞI UTILIZAREA GRAFICELOR 37
3.5. UTILIZAREA TABELELOR 39

CAPITOLUL 4. DETERMINAREA TENDINŢEI CENTRALE 43

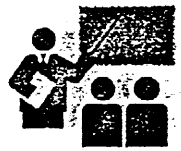
4.1. MEDIA ARITMETICĂ 43
4.1.1. Calculul mediei pe date negrupate. *44*
4.1.2. Calculul mediei unei distribuţii de frecvenţă *44*
4.1.3. Determinarea mediei utilizând valori codificate *47*
4.1.4. Combinarea mediilor provenite de la eşantioane diferite. *50*
4.1.5. Media procentajelor. *52*
4.2. MEDIANA 54
4.2.1. Mediana datelor negrupate. *54*
4.2.2. Mediana datelor grupate *55*
4.3 MODUL 58
4.4. MEDIA - MEDIANA - MODUL - CONSIDERAŢII FINALE - 60

CAPITOLUL 5 MĂSURI ALE DISPERSIEI 62

5.1. RANGUL. 63
5.2. ABATEREA MEDIE 64
5.3.ABATEREA STANDARD 65
5.3.1. Abaterea standard la eşantioane mici *67*
5.3.2. Abaterea standard la eşantioane mari *69*
5.3.3. Abaterea standard prin valori codate *70*
5.3.4. Corecţia lui Sheppard *71*

CAPITOLUL 6 CORELAŢIA 76
6.1. NOŢIUNEA DE CORELAŢIE 76
6.2. DIAGRAMA DE CORELAŢIE 77
6.3 CALCULUL COEFICIENTULUI DE CORELAŢIE 81
6.3.1. Coeficientul de corelaţie Bravais - Pearson (r). *81*
6.3.2. Coeficientul de corelaţie al rangurilor. *87*
6.4. INTERPRETAREA COEFICIENTULUI DE CORELAŢIE. 90
EXERCIŢII 94
BIBLIOGRAFIE SELECTIVĂ 100
ANEXA 1 VALORILE SEMNIFICATIVE ALE LUI *R* 101

ANDRIEŞ C. LUCIAN
Nr 2774
Biblioteca Personală



Introducere

În ultimii ani, metoda statistică s-a impus tot mai mult în ştiin umane, cunoscând o adevărată revoluţie în zilele noastre, prin penetra masivă a calculatoarelor. Astfel, analiza statistică a datelor a deveni activitate obişnuită, şi rapidă, specialiştii având la dispoziţie pachete programe statistice din ce în ce mai performante.

Obiectul principal al lucrării de faţă este familiarizarea studenţi din ştiinţele umane şi în special a celor din psihologie, sociologie pedagogie cu utilizarea metodei statistice. În acelaşi timp s-a avut în ved pregătirea în general umanistă a studenţilor de la aceste specialităţi, mo pentru care s-a renunţat la prezentarea detaliată a aparatului matematic.

De la început trebuie să menţionăm faptul că, statistica aplicată ştiinţele umane nu constituie o ramură a matematicii. Pentru studen practicienii şi cercetătorii din aceste ştiinţe, statistica aplicată reprezintă **mijloc** şi nu un scop.

DE CE ESTE IMPORTANTĂ STATISTICA?

Foarte pe scurt iată câteva motive:

1. *Este cel mai adecvat mod de descriere.*
Scopul final al oricărei cercetări este *descrierea* fenomenului studi Această descriere trebuie să fie în acelaşi timp completă, sintetică şi făc cu mare acurateţe. Doar în acest fel, datele vor fi înţelese şi devin u pentru cel ce citeşte simbolurile prin care fenomenul în cauză a fost desc Din acest punct de vedere, statistica aplicată este parte integrantă a aces limbaj descriptiv.

2. *Prin statistică suntem capabili să concentrăm rezultatele într formă convenabilă şi inteligibilă.* În urma cercetărilor efectuate, de cele m multe ori, suntem copleşiţi de un mare număr de date, rezultate

observaţii aflate într-o formă neinteligibilă. Numai statistica dispune de procedee şi mijloace adecvate pentru a face ordine in acest haos şi de a vizualiza forma şi tendinţa generală a rezultatelor.

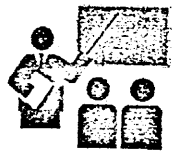
3. *Statistica permite conturarea concluziilor generale.* În procesul de extragere a concluziilor sunt obligatorii parcurgerea unor paşi. Pătrunzând înţelesul acestor paşi statistici vom putea aprecia *cât de mare poate fi încrederea acordată acestor concluzii.* De asemenea, vom fi în stare să apreciem şi gradul de generalizare şi extrapolare a rezultatelor la nivelul populaţiei în cauză.

4. *Utilizând statistica vom putea face predicţii, previziuni de genul "cât de mult".* Cu alte cuvinte, prin statistică suntem capabili să prevedem cât de mult se va adeveri dintr-un fenomen studiat în condiţii experimentale. Bineînţeles, pot fi şi erori cauzate de factori neprevăzuţi. Dar, tot statistica este aceea care ne va preciza care este marja de eroare a predicţiei în cauză.

5. *Nu în ultimul rând statistica permite efectuarea unor analize a factorilor complecşi ce pot denatura evenimentele.* Pentru ştiinţele educaţionale în general este caracteristic faptul că, concluziile obţinute sunt rezultatul a numeroşi factori cauzali. Factorii cauzali în mod obişnuit, cel mai bine sunt descoperiţi prin metoda experimentală, unde există posibiilitatea controlării tuturor factorilor implicaţi. Din păcate, acest lucru nu este întotdeauna posibil. În aceste situaţii metodele statistice pot într-o măsură oarecare să substituie un experiment, dar şi în aceste condiţii se pleacă de la date observate şi culese din realitate. Trebuie să subliniem faptul că, statistica şi experimentul întotdeauna merg mână în mână.

Având în vedere că principalii beneficiari ai lucrării de faţă sunt studenţii şi tinerii cercetători s-a încercat o cât mai mare apropiere de prelucrarea automată a datelor dar, accentuând în permanenţă modul de *interpretare* corectă a rezultatelor statistice. În acelaşi timp s-a avut în vedere şi explicarea fenomenelor statistice şi a motivelor pentru care a fost utilizată o tehnică statistică sau alta.

Capitolul 1 NOŢIUNI GENERALE



- **scor**
- **rang**
- **măsură**
- **scală** → **nominală**, **ordinală**, **de intervale**, **de proporţii**

Atunci când se studiază un proces psihic, o anumită trăsătură a personalităţii, datele obţinute în urma aplicării diferitelor metode de *măsurare* se prezintă adesea sub formă cantitativă sau numerică . În psihologie, sociologie şi pedagogie, aceste metode de măsură de cele mai multe ori sunt: teste, examene sau experimente.

Scorul . Numărul de puncte obţinut de un subiect la un test, examen, etc. este numit ***scor*** sau ***cotă***. De exemplu: subiectul V.V. a obţinut un punctaj de 32 la TAG. (Test General de Aptitudine). În această situaţie, scorul sau cota subiectului este: 32.
Un alt exemplu: subiectul G.G obţine la măsurarea timpului de reacţie mediu $t = 0{,}351$ sec. Cota sau scorul subiectului în acest caz este: 0,351.

Menţionăm că, în psihologie, sociologie şi pedagogie termenii de *scor* şi *cotă* sunt utilizaţi în egală măsură, ambele desemnând acelaşi lucru.

Subliniem faptul că, scorul sau cota nu indică locul, poziţia pe care-l ocupă subiectul în cauză în grupa din care face parte. De asemenea, scorul şi cota variază de exemplu, în funcţie de nivelul aptitudinal (în sensul că o aptitudine se poate perfecţiona trăgând după sine modificarea scorului). Această variaţie nu trebuie să fie în mod obligatoriu şi proporţională.

Rangul. În anumite situaţii suntem interesaţi de locul, respectiv *rangul* pe care-l ocupă un subiect într-o clasificare, făcută pe baza unui criteriu bine determinat. Astfel, se stabileşte care este primul, al doilea, al treilea ş.a.m.d., într-un colectiv dat putându-se efectua diferite comparaţii. Ca şi definiţie: *rangul este un număr care indică poziţia unei valori (al unui subiect) într-o serie ordonată (crescător sau descrescător) de valori (subiecţi) în funcţie de un criteriu apriori stabilit.*

De exemplu: într-un grup de subiecţi având ca şi criteriu înălţimea vom putea spune că subiectul A este cel mai înalt, iar cel mai mic este subiectul E. În acelaşi sens, vom putea spune că subiectul C este mai mare decât subiectul D, dar acest lucru nu informează cu nimic despre valoarea acestei diferenţe.

O altă modalitate, mai puţin utilizată în psihologie prin care putem imprima un anumit grad de obiectivitate materialului obţinut este folosirea *notaţiilor calitative.*

În astfel de situaţii, datele brute sunt grupate în clase valorice (de exemplu: bun, mediocru, slab). Scara calificativelor se poate nuanţa în funcţie de cerinţe sau scopul urmărit în 5, 7, 9 clase, ş.a.m.d.

Măsura. În sens larg, măsurarea este: *atribuirea de numere obiectelor sau evenimentelor în funcţie de anumite reguli bine determinate.* Dintre aceste reguli, două se impun cu prioritate: menţinerea constantă a unităţii de măsură de-a lungul întregii scale, iar seria valorilor posibile să înceapă de la zero absolut.

S.S. Stevens(1951) determină patru scale de măsură ce sunt frecvent utilizate în psihologie, fiecare având reguli proprii, posibilităţi statistice bine precizate şi anumite limite.

Scala nominală - prima şi cea mai slabă din seria de scale. Aceste *scale nominale servesc doar la denumire, identificare sau clasificare*, ele neavând proprietăţi matematice. Scala permite doar determinarea frecvenţei în fiecare categorie. Datele pot fi reprezentate grafic doar sub formă de histogramă.

Exemple: numerele de telefon; într-un grup de persoane când acordăm codul 1 pentru bărbaţi şi codul 2 pentru femei.

Scala ordinală - utilizată *când datele pot fi ierarhizate în funcţie de un criteriu* (crescător sau descrescător), dar nu putem determina exact diferenţa dintre două date. Reprezentarea grafică (histograma) poate lua forma de I, J sau clopot.

Aceste scale permit utilizarea procedurilor şi a tehnicilor statistice bazate pe utilizarea rangurilor.

Scala de intervale - unde se disting caracteristicile scalei ordinale, plus faptul că, de această dată sunt precis estimate distanţele dintre categorii/clase. Exemplu: majoritatea testelor vizând performanţele şcolare. De această dată pot fi calculate: media, ca şi tendinţă centrală şi abaterea standard ca şi indicator al dispersiei. Reprezentarea grafică a colecţiei de date se poate face sub formă de curbă.

Scala de proporţii - cel mai înalt nivel de măsurare. Aici, *fiecare număr reprezintă distanţa de la punctul zero.* De exemplu, prezentarea pe un liniar a distanţelor de 5 cm şi 11 cm. Ambele se fac din punctul zero al liniarului. La acest tip de scală sunt utilizabile toate operaţiile matematice şi statistice.

Numărul în măsurări. În măsurările psihologice, se impune o deosebită atenţie şi prudenţă în sensul că nu toate operaţiile cunoscute şi aplicabile numerelor se pot utiliza la scorurile obţinute. De exemplu: numărul=100. Acest număr are o **anumită semnificaţie** dacă se referă la nivelul intelectual al unui subiect şi este total diferit în cazul când se referă la greutatea subiectului.

Trebuie menţionat şi faptul că, *un număr oarecare reprezintă de fapt, un grupaj de valori cu o întindere de o unitate.* Astfel, acelaşi număr poate fi o valoare *minimă, maximă* sau *medie*. De exemplu: numărul 8 poate fi o valoare minimă în cazul lui Alex, care are 8 ani, pe care i-a depăşit şi se apropie de al 9-lea an. Acelaşi număr 8 poate fi o valoare maximă în cazul în care Alex a rezolvat corect toate cele 8 întrebări ale problemei la matematică. În mod obişnuit în statistica aplicată în psihologie, un scor este considerat ca o valoare medie, reprezentând un interval. De exemplu: scorul 8 reprezintă intervalul 7,5-8,5, scorul 19 reprezintă intervalul 18,5-19,5.

Având ca şi fundament cele mai sus menţionate despre măsurare şi considerarea numerelor în psihologie, putem începe incursiunea printre noţiunile statistice şi utilizarea lor într-un mod cât mai eficient.

ANDRIEŞ C. LUCIAN
Nr 2774
Biblioteca Personală

Capitolul 2 SOFTWARE STATISTIC

Odată cu evoluția tehnicii în domeniul calculatoarelor și a produselor software, a devenit posibilă analiza statistică a datelor și de către specialiștii din domeniul științelor sociale, datorită interfețelor prietenose, ușor de folosit ale numeroaselor pachete software existente.

În privința software-ului dezvoltat în științele sociale, există o mare varietate de programe atât cu scop general, cât și specializate pe anumite probleme. În cele ce urmează, vom încerca să facem o scurtă trecere în revistă a tipurilor de programe, insistând pe pachetele statistice și caracteristicile lor.

1. ***Generatoare de experiment în psihologie:*** au fost dezvoltate pentru a furniza cercetătorilor și studenților instrumente care să le permită efectuarea de studii psihologice, fără a investi timp și efort pentru a programa efectiv. Exemple: MEL, ERTS, SuperLab
2. ***Diagnosticare psihologică:*** programe de diagnosticare psihologică asistată de calculator ce conțin peste 100 de teste binecunoscute (matrici Raven, Cognitrone, etc.). Exemplu : Vienna Test System
3. ***Observații comportamentale și semnale fiziologice, analiza cardiovasculară:*** programe soft pentru înregistrarea de evenimente fiziologice sau semnale în vederea investigării anumitor caracteristici. Exemplu: CAMERA, CARSPAN, The Observer
4. ***Chestionare și teste***
   - 4.1. Bănci de itemi și software pentru construcția automată a testelor: The Examiner, ConTEST
   - 4.2. Intervievare asistată de calculator: programe ce permit proiectarea unui studiu, colectarea și editarea datelor, precum și pregătirea datelor pentru prelucrarea statistică. Blaise III permite în plus intervievarea asistată de calculator prin telefon, personal sau autointervievarea: Exemple: Blaise III, Edwin 2

5. ***Cercetare calitativă***: programe dezvoltate pentru analiza de text, aceasta permițând analize cantitative pe baza unor texte. AQUAD 5, AGREE 6 , etc
6. ***Design de experimente***: programe ce permit planificarea optimă a experimentelor și identificarea modelelor statistice care descriu cel mai bine setul experimental de date. CADEMO 3, RANCODE-Plus
7. ***Analiza rețelelor*** sociale: măsoară centralismul și conectivitatea, determină subgrupuri și poziția în grup, etc. UCINET, GRADAP 2.
8. ***Modelarea ecuațiilor structurale:*** estimează RELații Structurale LIniare recursive și nerecursive (LISREL). Amos 3.6 permite modelare structurală, analiza structurilor de covarianță și a modelelor cauzale, acceptă diagrame path pentru specificarea modelului.
9. ***Analiza multinivel***: analizează modele cu mai multe nivele de imbricare: HLM 4, MLn
10. ***Scalare***: programe bazate pe teoria răspunsului la itemi (item response theory): BILOG, SCA, MULTILOG 6, PARSCALE 3, SIBTEST, etc
    - 10.1. Analiză de itemi: realizează analiză de itemi, analiza factorială a itemilor, atât a datelor dihotomice , cât și politomice. Testfact 2, ITEMAN 3
11. ***Sisteme Expert***: sistem de ghidare inteligentă în procesul de analiză statistică: SigmaStat 2.0

12. ***Pachete statistice cu scop general:***

11.1. ***SPSS - pachetul de bază:*** oferă facilități de calcule statistice, grafice și rapoarte. Deasemenea permite importul, respectiv exportul de date, facilități de gestionare a variabilelor (recodare, selecție, sortare, calcul de noi variabile). Procedurile statistice pornesc cu statisticile descriptive, tabele de contingență, și continuă cu analiza de varianță, regresia și teste neparametrice.

11.2. ***SPSS - module***

11.2.1. *SPSS Professional statistics:* realizează analiză cluster, discriminant, factorială, plus tehnici de regresie avansată ca tehnica celor mai mici pătrate ponderate și cele mai mici pătrate în doi pași

11.2.2. *SPSS Advanced statistics:* include tehnici univariate și multivariate, proceduri ale modelului general liniar, MANOVA, estimări Kaplan-

Meier, regresie Cox, regresie neliniară, analiză probit şi regresie logistică

11.2.3. *SPSS Tables*: crează o varietate de prezentări tabelare utile în rapoarte

11.2.4. *SPSS Categories:* analizează datele categoriale, realizează analiză conjoint şi scalare optimală, incluzând analiza de corespondenţă

11.2.5. *SPSS Trends* efectuează analiza seriilor de timp şi conţine metode pentru estimarea funcţiilor autoregresive

11.2.6. *SPSS Chaid* dezvoltă modele predictive

11.2.7. *SPSS Exact tests* este potrivit pentru efectuarea de teste statistice când se folosesc seturi mici sau debalansate de date. *Exact Tests* includ o serie de algoritmi pentru calcularea valorilor p şi a intervalelor de confidenţă pentru un interval larg de teste pe tabele de contingenţă şi date neparametrice.

11.3 ***SYSTAT***: oferă facilităţi de lucru prin grafică interactivă, editare de tip spreadsheed, utilizatorii avansaţi putând executa analiză multivariată de varianţă şi regresie neliniară. Mai oferă multiple metode de analiză cluster, analiză factorială incluzând analiza în componentele principale şi modele factor obişnuite, scalare multimensională, statistici neparametrice, corelaţii canonice, etc. Desemenea, SYSTAT poate estima şi testa orice model liniar univariat sau multivariat.

11.4. ***Statistica***: include o selecţie bogată de statistici şi posibilităţi grafice într-un singur sistem.

11.5. ***S-Plus*** este un instrument extensibil şi cu multiple posibilităţi de modificare din partea utilizatorului, util atât analizelor exploratorii cât şi tehnicilor statistice. Fiind un sistem orientat obiect, S-Plus mânuieşte toate datele, funcţiile şi modelele ca obiecte, făcând analizele mult mai flexibile decât o abordare de tip procedural. Programul compară, şi validează statistic modelele alternative, rezultatele fiind imediat disponibile ca obiecte în analize viitoare. Desemenea, programul poate interacţiona cu alte limbaje ca C, Fortran şi Mathematica.

11.6. ***Minitab***, ca şi celelalte programe văzute până acum, permite efectuarea de statistici de bază, regresii, ANOVA, descrierea de experimente, trasarea de grafice, analiza seriilor de timp, teste

neparametrice, simulări şi distribuţii. În plus, Minitab suportă conexiunile de tip ODBC, ceea ce permite utilizatorului să preia date dintr-un pachet ca Access, Oracle, Sysbase sau Informix. Deasemenea, macrourile Minitab permit utilizatorului să-şi creeze propriile sale operaţii.

11.7. ***NCSS 6.0***: program dezvoltat în special pentru nestatisticieni care au nevoie de o analiză inteligentă a datelor şi de a scrie uşor rapoarte bazate pe rezultatele găsite

11.8. ***ViSta*** , (The Visual Statistics System), software free accesibil pe Internet la adresa http://forrest.psych.unc.edu cu facilităţi grafice.

Lista ar putea fi mult mai lungă şi mai detaliată; ceea ce s-a urmărit însă nu este discutarea comparativă a software-urilor existente, ci aducerea la cunoştinţa celor interesaţi de multitudinea de oferte în ceea ce priveşte software-ul din ştiinţele sociale. Aceste programe sunt în continuă perfecţionare, astfel încât este necesară în primul rând o cunoaştere foarte aprofundată a domeniului în care ele se aplică (de exemplu conceptele statistice, modalităţile de generare de experimente în domeniul ştiinţelor sociale, etc.). Simpla aplicare a procedurilor acestor programe, fără o temeinică cunoaştere teoretică a conceptelor folosite, nu va putea niciodată să ducă la o cercetare valoroasă.

Ca şi concluzie, utilizatorul este invitat să parcurgă alături de noi exemplificările făcute cu ***SPSS Pachetul de bază şi module,*** versiunea *6.1*. Alegerea acestui pachet s-a făcut datorită relativei răspândiri mai mari a acestui program ca utilizare. Însuşirea conceptelor de bază ale utilizării acestui program va face posibilă folosirea şi a altor programe de calcul statistic, decizia de folosire rămânând în sarcina utilizatorului, conform cerinţelor lui proprii.

2.1. SPSS - Statistical Package for Social Science

În acest capitol vom aminti foarte sumar câteva din etapele realizării unui studiu în ştiinţele sociale, deoarece considerăm util acest lucru în demersul nostru privind descrierea anumitor concepte ca *date, variabile, înregistrări*, etc., elemente de bază manipulate de SPSS.

- versiunea cu care lucrăm este: SPSS 6.1 pentru Windows 3.11, (putându-se folosi şi sub Windows 95)
- ocupă ≈ 20Mb instalat, iar sub formă de kit de instalare este disponibil pe 12 dischete.
- informaţii referitoare la SPSS sunt valabile la adresa http://www.spss.com

Cum se prezintă pachetul software SPSS:

- este sistem bazat pe ferestre ce posedă patru ferestre principale:
  1. fereastra principală a programului SPSS
  2. fereastra de date
  3. fereastra rezultatelor (output)
  4. fereastra de sintaxă (pentru programarea în limbajul de comenzi SPSS)

2.1.1 Primii paşi

Ce trebuie să faci pentru a explora o idee?

- să-ţi formulezi cu atenţie problema şi să decizi care sunt informaţiile necesare pentru a răspunde la problemă
- trebuie să determini care este populaţia de interes şi să selectezi un eşantion aleator de obiecte sau persoane ale acelei populaţii
- trebuie să fii sigur că în cadrul eşantionului probabilitatea de selecţie a unor membrii nu este mai mare decât a altor membrii
- trebuie să-ţi colectezi informaţiile într-o manieră obiectivă. Procedurile de achiziţionare a informaţiilor trebuie să fie obiective şi standardizate. Întrebările trebuie să fie neambigue.
- trebuie să previi subiecţii şi investigatorii să nu influenţeze personal investigaţia.

2.1.2.Proiectarea unui studiu

Când proiectăm un studiu, trebuie să avem în vedere:

- ce fel de informaţii vom achiziţiona de la fiecare subiect
- cum vom înregistra răspunsurile
- cum asignăm coduri la răspunsuri

- cum vom indica că un subiect refuză să răspundă sau că informaţia nu este valabilă
- cum ne pregătim chestionarul pentru ca ulterior datele să fie introduse cu uşurinţă în calculator

Deoarece este mult mai uşor să urmărim explicaţiile axaţi pe un exemplu concret, vom folosi un studiu pe scară-largă, numit **General Social Survey, GSS,** efectuat la National Opinion Research Center, o organizaţie de cercetări în ştiinţe sociale de la Universitatea din Chicago.GSS este administrat în fiecare an unui lot de 1500 persoane, eşantion ce reprezintă populaţia adultă din SUA. Un intervievator chestionează o persoană aleasă, **subiectul**, asupra experienţelor prezente şi trecute, comportament, opinii.

Metoda de selecţie a persoanelor pentru a fi incluse în eşantion este complexă. Implică în primul rând o selecţie a unui eşantion aleator de oraşe şi judeţe, apoi o selecţie a unui eşantion aleator de vecinătăţi, apoi o selecţie a unui eşantion aleator de locuinţe şi în sfârşit o persoană din locuinţă.

GSS a fost proiectat să includă întrebări variate, şi deasemenea întrebări puse în cadrul unor studii din anii 1940, 1950, 1960, etc. Având datele de la aceleaşi întrebări puse în decursul mai multor ani, se poate studia cum s-au schimbat atitudinile şi opiniile.

Proiectarea chestionarului

Exemplul 1: O selecție de întrebări ale chestionarului pentru GSS

****

1. Număr înregistrare ☐☐☐☐

2. Data nașterii ☐☐ ☐☐ ☐☐☐☐

Lună Zi An

Notațiile numerice ale lunilor

Ianuarie	01	Mai	05	Septembrie	09
Februarie	02	Iunie	06	Octombrie	10
Martie	03	Iulie	07	Noiembrie	11
Aprilie	04	August	08	Decembrie	12

3. Sex:

Masculin.....1 Feminin.....2

4. Stare civilă

Căsătorit.....1 Separat.....4

Văduv.....2 Necăsătorit.....5

Divorțat.....3

****

Codificarea datelor:

• în cazul în care anumite răspunsuri la întrebări sunt numere, se va lăsa spațiu pentru completare

• dacă răspunsul nu este un număr, se va încerca în avans să se prevadă eventualele răspunsuri, subiectul urmând să aleagă între alternative

• dacă nu se știe în avans ce vor răspunde subiecții și ne interesează exact ceea ce spun ei, atunci avem de-a face cu așa-numitele **întrebări deschise**; în acest caz vom asigna coduri la aceste răspunsuri pentru a putea fi introduse în calculator

• prevedeți alternative de răspuns gen: "Nu știu", "Asta nu-i treaba ta!"

Deci: Cum proiectați chestionarul pentru a putea fi folosit ulterior la înregistrarea datelor?

• indicați un loc specific în chestionar pentru înregistrarea fiecărei informații
• includeți un număr de identificare pentru fiecare subiect
• înregistrați valorile actuale pentru datele numerice
• dacă este necesar, asignați coduri la posibilele răspunsuri
• asigurați-vă că datele vor putea fi introduse în calculator direct din chestionar

2.1.3. Date în SPSS

În cele ce urmează, vom folosi **General Social Survey,** în care selecția eșantionului, proiectarea formei studiului, antrenarea operatorilor de teren, aranjarea datelor și introducerea acestora pe calculator au fost executate de profesioniști experimentați. Fișierul în care au fost înmagazinate aceste date a fost denumit **gss91t.sav, .sav** fiind extensia implicită pe care o atașează SPSS-ul fișierelor de date.

2.1.3.1. Pregătirea introducerii datelor

1. Se codifică întrebările chestionarului. Pentru situațiile de nonrăspuns, lipsa răspunsurilor, etc. se vor prevedea coduri speciale.
2. Se verifică toate chestionarele pentru a lua în considerare eventualele situații neprevăzute și a fi codificate.
3. Se introduc datele pe calculator.

2.1.3.2 Introducerea datelor

Datele se pot introduce:

1. în fereastra de date a programului SPSS
2. în alte medii de gestiune a datelor: Excel, Lotus, SYLK, dBASE, FoxPro
3. în versiunea DOS a SPSS-ului, numită SPSS/PC+

4. în mod text, câmpurile fiind delimitate de tab-uri

Pentru datele introduse cu variantele 2, 3, 4, SPSS-ul va realiza un *import* de date.

2.1.3.3. Definirea datelor

• *înregistrări, variabile, valori, etichete*

Înregistrare (case): fiecare subiect al unui studiu sau experiment (observații).

Variabilă (variable): tipul de informație pentru care se înregistrează răspunsuri (data nașterii,sex, educație,etc.).

Valoare (value): răspunsul fiecărui subiect la variabilă.

Eticheta variabilei (variable label): descrierea pe larg a fiecărei variabile, în cuvinte.

Eticheta valorii (value label): descrierea pe larg a valorilor cu o anumită semnificație (coduri), în cuvinte.

Valori lipsă (missing values): fiecărei variabile considerate i se va pune în corespondență un cod de NONRĂPUNS în cazul în care valorile acestei variabile lipsesc, sau au fost introduse greșit, etc.

Exemplul 2: Să ne definim elementele unui studiu social general propriu.

Variabile: cod_id, ani, sex, stareciv, satjob, fercas, viata, postviata, educ, studii, taeduc, maeduc, buget, sanat, nrpers

Etichetele variabilelor:

cod_id:	"Cod de identificare"
ani:	"Vârsta subiectului"
sex:	"Sexul subiectului"
stareciv:	"Stare civilă"
satjob:	"Satisfacția muncii"
fercas:	"Fericirea în căsnicie"
viața:	"Viața este frumoasă sau plicticoasă?"
postviața:	"Credința în viața de după moarte"
educ:	"Educația"

studii:	"Diploma - calificarea"
taeduc:	"Educația tatălui"
maeduc:	"Educația mamei"
buget:	"Câștigul total al familiei"
sanat:	"Starea sănătății"
nrpers:	"Numărul de persoane din casă"

Etichetele valorilor:

sex	1 - masculin 2 - feminin
satjob	1 - foarte satisfăcut 2 - moderat 3 - nesatisfăcut 4 - foarte nesatisfăcut
fercas	1 - foarte fericit 2 - fericit 3 - nu prea fericit 9 - nonrăspus
postviața	1 - da 2 - nu 9 - nonrăspus
studii	0 - studii universitare 1 - colegiu 2 - liceu 3 - școală profesională 4 - școală generală 9 - nonrăspuns
ani, educ, taeduc, maeduc	99 - nonrăspuns
cas	1 - căsătorit 2 - necăsătorit 3 - văduv 4 - divorțat 9 - nonrăspuns
viata	1 - palpitantă 2 - rutinantă 3 - plicticoasă 9 - nonrăspus
buget	1 - sub 100.000 2 - 100.000 - 250.000 3 - 250.000 - 400.000 4 - 400.000 - 600.000 5 - peste 600.000 99 - nonrăspuns
sanat	1 - excelentă 2 - bună 3 - acceptabilă 4 - proastă 9 - nonrăspuns

Observație: Variabilele *ani, educ, taeduc, maeduc* sunt numerice (vezi Capitolul 1, la scala de interval și de proporții), astfel încât ele nu vor avea răspunsuri predefinite; ceea ce putem noi să specificăm este doar codul 99 pentru cazul de NONRĂSPUNS.

2.1.3.4. Deschidere şi salvare fişiere SPSS. Import fişiere în SPSS.

Notă: pe tot parcursul exemplelor practice în SPSS vom folosi următoarele notaţii: M → Meniu, O → Opţiune, SubO → Subopţiune.

1. ***Crearea fişierelor noi***: M **File** O **New**:
   - SubO **Data**: crează fişiere noi de date SPSS (.sav)
   - SubO SPSS **Syntax**: crează fişiere noi care conţin programe în limbajul de comenzi al SPSS-ului (.sps)
   - SubO SPSS **Output**: crează fişiere noi care conţin rezultate ale unor comenzi SPSS (.lst)

2. ***Deschiderea fişierelor existente*** : M **File** O **Open**:
   - SubO **Data**: deschide fişiere de date SPSS (.sav)
   - SubO **Chart**: deschide fişiere care conţin chart-uri (diagrame grafice) (.cht)
   - SubO SPSS **Syntax**: deschide fişiere care conţin programe în limbajul de comenzi al SPSS-ului (.sps)
   - SubO SPSS **Output**: deschide fişiere care conţin rezultate ale unor comenzi SPSS (.lst)

3. ***Închiderea ferestrei curente***: M **File**, O **Close**
4. ***Salvarea fişierelor***
   - salvarea fişierelor de date (cu înregistrări), de tip .sav : M **File** O **Save Data**
   - salvare fişiere de mai multe tipuri: M **File** O **Save As...**
     - fişiere de date SPSS: (.sav)
     - fişiere de date SPSS/PC+: (.sys)
     - fişiere de date SPSS Portable: (.por)
     - fişiere de date - dBASE, FoxPro (.dbf), Excel (xls), Lotus (.w*),
     - fişiere de date delimitate cu tab-uri: (.dat)

5. ***Tipărirea la imprimantă***: M **File**, O **Print**
   - configurare imprimantă: M **File**, O **Printer Setup**

6. ***Oprirea execuţiei***: M **File** O **Stop SPSS Processor**
-această comandă se foloseşte în momentul în care este activă o comandă SPSS şi dorim să o întrerupem.
7. ***Părăsire SPSS***: M **File** O **Exit**

Cum importăm un fişier de date ?

Când dorim să lucrăm cu un fişier de date (fişier bază de date ce conţine înregistrări), creat cu un alt program (dBASE, FoxPro, etc.), executăm următoarele:

- din M **File** O **Open**, alegem tipul de fişier care dorim să-l importăm .dbf, .xls, etc. selectând tipul dorit din lista **File Type**; apoi din lista fişierelor existente de acel tip, ne alegem fişierul dorit.
- după ce fişierul a fost deschis, îl vom putea salva cu ajutorul M **File** O **Save As...**cu extensia .sav, fişierul putând fi prelucrat mai departe în SPSS.

Capitolul 3 STATISTICI DESCRIPTIVE

• **colecție de date**

În urma unei investigații psihologice datele obținute se prezintă de cele mai multe ori sub formă numerică sau sunt aduse în această formă grație operațiilor de codare. Aflate sub această formă datele se pretează la prelucrare statistică.

Trebuie remarcat că, aceste date cantitative, scoruri brute cu care operează psihologul, rezultă pe două căi:

- în mod *direct* - citirea directă a aparatelor (cronometru, contor, număr erori etc.)
- în mod *indirect* - rezultat prin însumarea, calcularea cotei brute, cum este cazul testelor creion - hârtie (baraj, nivel intelectual, etc.), sau totalizări de rubrici ordonate (protocol de observație).

Aceste colecții de date brute, așa cum au fost ele obținute nu ne spun, nu ne dezvăluie nimic, practic sunt *ansambluri aleatoare* de date (Tabelul 2.1). Numai după ordonarea și gruparea datelor utilizând tehnicile *statisticii descriptive* vor fi dezvăluite elementele, tendințele esențiale ale acestor date.

Exemplu: Un psiholog școlar a aplicat la un număr de 72 elevi din clasa a XII-a un test de memorie, mai exact de evaluare a capacității de memorare. Cotele brute obținute au oscilat între 54-104 cuvinte; cu cât cota este mai mare cu atât mai eficace este capacitatea de memorare (Tabelul 3.1).

Tabelul 3.1 Rezultatele obținute de cei 72 elevi la proba de memorare

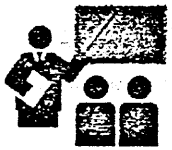
84	95	71	60	54	76	87	82
89	82	91	85	92	84	69	81
74	90	104	80	85	81	74	
79	85	71	76	59	68	75	
81	77	78	73	76	83	79	
66	74	88	94	61	79	77	
98	72	89	72	65	66	83	
76	64	67	79	87	70	78	
70	90	86	87	83	82	84	
68	93	92	96	80	73	89	

După cum se poate vedea din Tabelul 3.1, în urma testării s-a obținut o **colecție de date** dispuse într-o *ordine aleatoare*. Pe noi ne interesează ne interesează pe noi ca din aceste date să putem extrage concluzii referitoare la capacitatea de memorare, scopul final urmărit de psihologul școlar în cauză.

Pentru acest lucru facem uz de statistica descriptivă, în primul rând, operând o ierarhizare, apoi o grupare a datelor și în final o prezentare grafică a acestor operații ce vor permite extragerea primelor concluzii.

Exemplul 3.1.: Crearea unui fișier de date cu rezultatele testului de memorie din Tabelul 3.1.

1. Se lansează programul SPSS.
2. Se crează variabila MEMORIE din M **Data**, O **Define Variable**, se completează câmpul **Variable Name** cu textul MEMORIE și se confirmă operația prin **OK.**
3. În coloana variabilei MEMORIE se înscriu una sub alta valorile testului de memorie.
4. Se salvează fișierul sub numele de testmem.sav din M **File**, O **Save As...**, completându-se câmpul **File Name** cu testmem.sav, iar operația se confirmă cu **OK.**



3.1. *Gruparea datelor*

- **clasă**
- **distribuție de frecvență**

O primă posibilitate de organizare a datelor este ierarhizarea, respectiv ordonarea după *rang* a acestora. Astfel, vom putea mult mai ușor să vizualizăm limitele și numărul de valori ale ansamblului de date, precum și frecvența de apariție a acestora. În acest sens Tabelul 3.1 s-a transformat în Tabelul 3.2.

Tabelul 3.2. Distribuția după rang a datelor din Tabelul 3.1.

104	90	85	82	78	74	69	59
98	89	85	81	78	74	68	54
96	89	84	81	77	73	68	
95	89	84	81	77	73	68	
94	88	84	80	76	72	66	
93	87	83	80	76	71	65	
92	87	83	79	76	71	65	
92	86	83	79	76	71	64	
91	86	82	79	75	70	61	
90	85	82	79	74	70	60	

Acest nou Tabel 3.2. vizualizează foarte clar limitele, valorile extreme, adică cota brută minimă și maximă obținută la testul de memorie. Mai remarcăm faptul că, unele valori sunt singulare, altele apar de mai multe ori.

Modul de distribuție după rang nu este operant în toate cazurile. În numeroase situații, psihologii și sociologii operează cu sute de date, situație în care metoda rangurilor devine greoaie și chiar inoperantă. În astfel de situații se trece la ***gruparea datelor***, într-un anumit număr de **clase**, dar care obligatoriu trebuie să fie *egale ca dimensiune*. În exemplul nostru putem

grupa datele în clase de câte cinci unități fiecare, adică vom utiliza un ***interval de grupare*** (i) egal cu cinci (i = 5).

Adică vom scrie:

104	103	102	101	100	ceea ce se scrie →	104-100
99	98	97	96	95		99-95
94	93	92	91	90		94-90
etc.						etc.

Utilizând metoda de mai sus, datele noastre (Tabelul 3.1) vor arăta astfel:

Tabelul 3.3. Gruparea datelor în clase și distribuția de frecvențe:

Cotele brute X	**Valoarea centrală** X_k	**Frecvența** f
104-100	102	1
99-95	97	3
94-90	92	7
89-85	87	11
84-80	82	14
79-75	77	13
74-70	72	11
69-65	67	7
64-60	62	3
59-55	57	1
54-50	52	1
i=5		N=72

Atragem atenția asupra faptului că, se impune o precauție în alegerea intervalului de frecvență și a numărului de clase, deoarece acestea pot influența *tendința de ansamblu* a fenomenului studiat. Subliniem că, doar studiul calitativ al ansamblului de date ne va arăta ce decizie să luăm cu privire la numărul de clase și mărimea intervalului cu care urmează să operăm.

Ceea ce s-a obținut în Tabelul 3.3 este o **distribuție de frecvență** sau **distribuție statistică** de la care pleacă orice prelucrare statistică a datelor obținute pe cale experimentală sau în urma unei testări. Aceeași formă primară de ordonare și grupare a datelor poate oferi și unele concluzii referitoare la **tendința** care se remarcă în ansamblul de date studiat.

Exemplul 3.2: Să grupăm datele din Tabelul 3.1. în 11 clase, a câte 5 membrii fiecare, folosind fişierul testmem.sav.

1. Vom grupa datele, folosind operaţia de **Recode** (recodificare) din M **Transform**, O **Recode**, SubO **Into Different Variable.**
2. Din lista de variabile, vom trece variabila MEMORIE în lista de transformare **Input Variable→Output Variable.**
3. În câmpul **Name** din **Output Variable** vom scrie numele variabilei de grupare, GRUPMEM, iar apoi vom face efectiv schimbarea de nume cu **Change.**
4. Apăsăm butonul **Old and New Values...**pentru a defini grupurile; apoi, în partea stângă, **Old Values**, alegem opţiunea **Range** (interval) unde vom înscrie prima grupă, 50 **through** 54.
5. În zona **New Value**, câmpul **Value** îl completăm cu valoarea 1 (indexul primei grupe).
6. Apăsăm butonul **Add** pentru a insera în lista **Old→New** prima grupă creată.
7. Continuăm în acelaşi mod înscrierea celorlalte grupe (55 **through** 59), apoi (60 **through** 64), etc.
8. După înscrierea tuturor grupelor, se apasă **Continue** în fereastra de **Recode into Different Variables: Old and New Values.**
9. Se confirmă operaţia prin **OK** în fereastra **Recode into Different Variables.**

În acest mod, noua variabilă GRUPMEM va conţine codurile de la 1 la 11, fiind stabilită astfel apartenenţa fiecărui caz la grupa corespunzătoare.

3.2. Reprezentări grafice

- **diagrama**
- **histograma**
- **curba frecvenţelor cumulate**
- **curba procentajelor cumulate**

O metodă foarte frecventă de vizualizare a datelor colectate este *prezentarea grafică.* Metoda este des utilizată datorită modului intuitiv prin care dezvăluie esenţa fenomenului aflat în studiu.

Metodele cele mai des utilizate în reprezentarea grafică sunt: ***diagramele*** (circulare şi de comparaţie), ***histogramele, poligonul frecvenţelor, curba frecvenţelor cumulate*** şi ***curba procentajelor cumulate.***

3.2.1 Diagramele

În practica psihologică, sociologică şi pedagogică se utilizează mai des două tipuri de diagrame în vederea vizualizării datelor şi anume: ***diagramele circulare*** şi ***diagramele de comparaţie.***

2.2.1.1. Diagramele circulare

La construirea acestor diagrame se utilizează cercul care se împarte în sectoare proporţionale cu valorile pe care le reprezintă. Notăm că, pentru acest lucru se impune ca valorile brute, cotele să fie transformate în *indici procentuali.*

Exemplu: Într-o cercetare privind capacitatea de decizie, un număr de 112 conducători auto au primit următoarele calificative în funcţie de tipul de decizie manifestat:

a) extrem de imprudenți -5 subiecți
b) imprudenți -11 subiecți
c) echilibrați -57 subiecți
d) prudenți -26 subiecți
e) excesiv de prudenți -13 subiecți

Aceste rezultate pot fi prezentate printr-o astfel de diagramă circulară sau de structură. Iată şi procedura de urmat:

1. Cercul are 360 de grade, deci 1/100 este egal cu 3,6° ;
2. Convertirea valorilor brute în indici procentuali. Acest lucru se poate realiza foarte simplu prin utilizarea regulii de trei simplă:

112 subiecți----------100%
5 subiecți ----------x

$$x = \frac{5 \times 100}{112} = 4{,}46(4\%)$$

112 subiecți----------100%
26 subiecți----------x

$$x = \frac{25 \times 100}{112} = 23{,}2(23\%)$$

La fel se procedează şi cu celelalte date.

3. Determinarea sectoarelor proporționale de cerc, corespunzătoare indicilor procentuali, folosind aceeaşi regulă de trei simplă:

360----------100% 360----------100%
x----------4% x----------23%

$$x = \frac{360 \times 4}{100} = 14{,}4(14) \qquad x = \frac{360 \times 23}{100} = 82{,}80(83)$$

În acelaşi fel se procedează şi pentru celelalte valori.

4. Construirea practică a diagramei este prezentată în Figura 3.1.

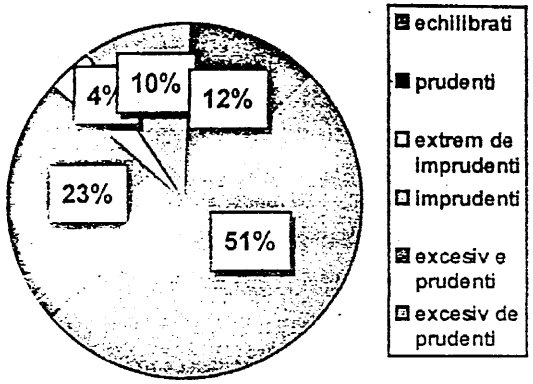


Fig.3.1.Diagrama de structură a repartiției calificativelor capacității de decizie.

3.2.1.2. Diagrama de comparație

Aceleaşi rezultate mai sus arătate pot fi reprezentate şi pe o diagramă de comparație. La aceste reprezentări pot fi utilizate atât procente cât şi frecvențe absolute. În caz de nevoie, se poate include şi o legendă cu explicații suplimentare.

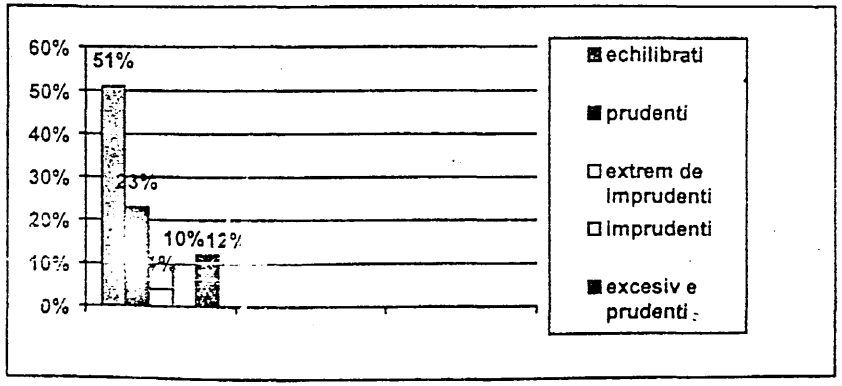


Fig. 3.2. Diagrama de comparație a calificativelor modalităților de decizie la conducătorii auto.

3.2.2. Histograma

O modalitate foarte des utilizată în prezentarea vizualizată a datelor experimentale este ***histograma***. Esența constă în utilizarea unor arii dreptunghiulare, sau sub formă de bare, folosindu-se sistemul axelor rectangulare. Astfel, pe axa absciselor (x) se marchează valorile brute, iar pe cea a ordonatei (y) vor fi trecute frecvențele absolute sau procentele.

Exemplu: La terminarea unui curs de pregătire la matematică în vederea prezentării la bacalaureat s-a aplicat un test de cunoştinţe la cei 110 cursanţi. Rezultatele au fost cuprinse între: 26 şi 102 răspunsuri corecte. Distribuția de frecvențe obținută este redată în Tabelul 3.4.

Tabelul 3.4. Distribuția de frecvență a rezultatelor la testul de cunoştinţe de matematică.

Cote Brute X	Valoarea Centrală X_k	Frecvența f
102-96	99	1
95-89	92	2
88-82	85	3
81-75	78	8
74-78	71	17
67-61	64	14
60-54	57	34
53-47	50	16
46-40	43	8
39-33	36	6
32-26	29	1
i=7		N=110

Iată şi procedeul practic de construire al histogramei :

1. Trasăm axele rectangulare (0x, 0y).
2. Pe axa ordonată (0x) marcăm 11 segmente egale, corespunzătoare celor 11 intervale de clasă. Notăm pentru fiecare clasă limitele aferente.

3. Pe axa abscisă (0y) trecem frecvențele absolute.
4. Trasăm ariile sub formă de dreptunghiuri corespunzătoare fiecărei clase.

În Figura 3.3. sunt redate sub formă de histogramă datele din Tabelul 3.4.

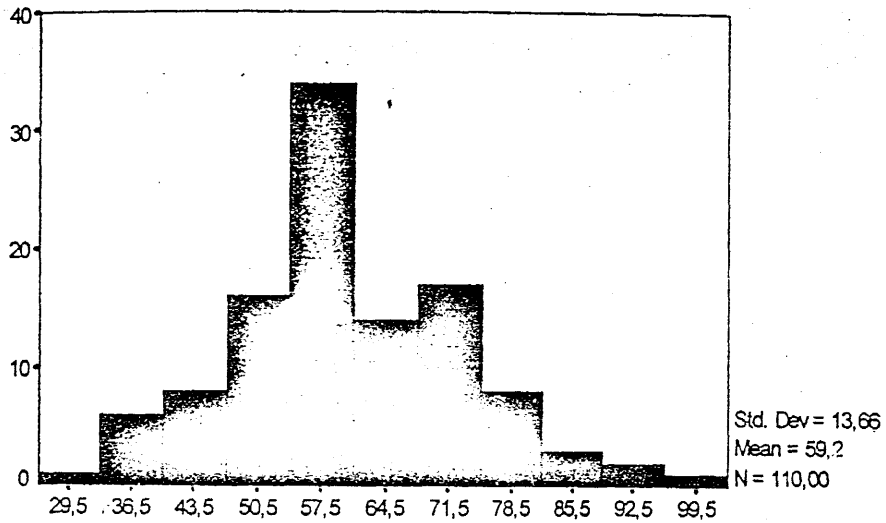
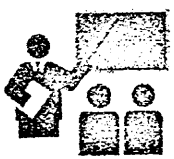


Fig. 3.3. Histograma distribuției de frecvență a rezultatelor obținute de cei 110 elevi la testul de cunoştinţe la matematică.

Exemplul 3.3 : Să trasăm histograma rezultatelor testului de memorie care au distribuția de frecvență în Tabelul 3.4

1. Se construieşte variabila SCORMATE, introducând cazurile în următorul mod: din fiecare grupă se vor introduce atâtea scoruri cât este frecvența grupei (de exemplu, din grupa 81-75 se vor introduce 8 cazuri: 75, 75, 76, 77, 78, 78, 78, 79).
2. Se recodifică SCORMATE în GRUPMATE aşa cum s-a arătat în Exemplul 3.2.
3. Se trasează histograma din M **Graphs**, O **Histogram**...prin includerea în câmpul **Variable** a variabilei SCORMATE.

4. În fereastra nou obținută ***Chart Carousel***, se apasă butonul **Edit** pentru a putea ajusta intervalul de grupare; cu mouse-ul se face un dublu click pe scala abscisei, în acest mod apărând pe ecran fereastra **Interval Axis.**
5. În zona **Intervals** se apasă butonul radio **Custom**; apoi se apasă butonul **Define...**, iar în fereastra nou obținută **Interval Axis: Define Custom Intervals** se stabilește din zona **Definition, Interval Width** (lățimea intervalului din histogramă) de 7, și din zona **Range, Minimum** de 26 și **Maximum** de 102.
6. Se apasă **Continue** în fereastra **Interval Axis: Define Custom Interval,** iar apoi **OK** în cea de **Interval Axis.**
7. Opțional se poate salva fișierul Chart (adică histograma obținută) prin M **File,** O **Save As...** cu numele scormate.cht.
8. Revenim în fereastra de date cu ajutorul M **Window**, alegând dintre ferestrele existente (*1!Output, 2Newdata, 3!Chart*) fereastra *Newdata*. Se salvează fișierul de date cu numele mate.sav (M **File** O **Save As...**),



3.2.3. Poligonul frecvențelor

O altă modalitate de prezentare a datelor experimentale sau rezultate în urma unei testări este ***poligonul frecvențelor***. În acest caz, pe axa ordonată se vor trece valorile centrale (X_k) ale claselor de valori, iar pe abscisă frecvențele absolute (ca și în cazul histogramei).

În Figura 3.4. este prezentat poligonul frecvențelor datelor prezentate în Tabelul 3.4.

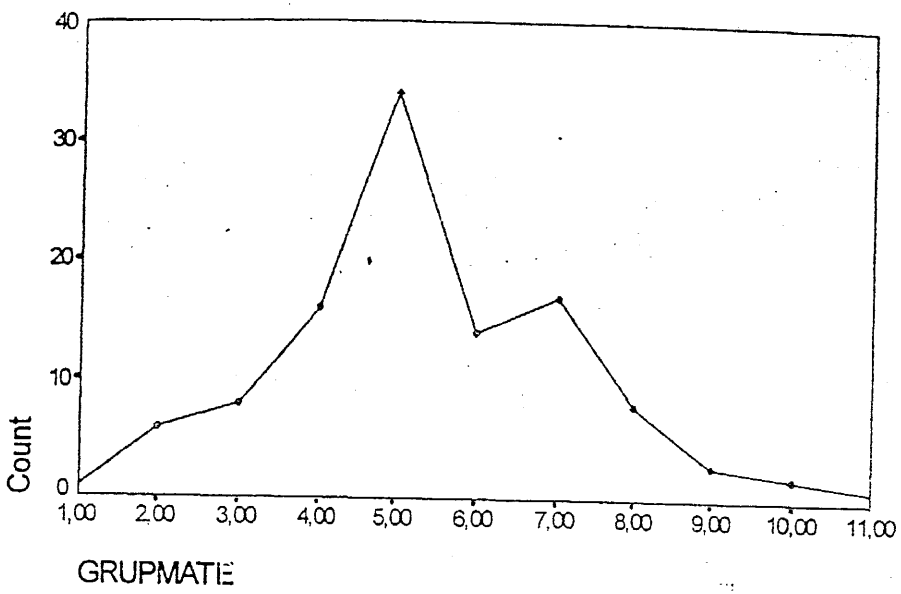
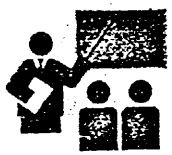


Fig. 3.4. Poligonul frecvențelor rezultatelor la testul de matematică.

Exemplul 3.4. Să trasăm poligonul frecvențelor rezultatelor la testul de matematică existente în fișierul mate.sav (vezi Exemplul 3.3).

1. Din M **Graphs** se alege O **Line...**
2. În fereastra de **Line Charts** se alege modalitatea **Simple.**
3. Se apasă butonul **Define**; câmpul **Category Axis** se completează cu variabila GRUPMATE.
4. Se confirmă acțiunea prin **OK.**
5. Opțional, se poate salva graficul cu M **File,** O **Save As...,** dându-se numele frecv.cht.



3.2.4. Curba frecvenţelor cumulate

Sunt foarte frecvente situaţiile în care se cere prezentarea grafică comparativă a două sau mai multe grupuri *inegale numeric*. În aceste situaţii, histograma sau poligonul frecvenţelor sunt neadecvate, cu alte cuvinte în aceste situaţii utilizarea cotelor brute pe abscisă este eronată. *În aceste cazuri se face apel la exprimarea frecvenţelor în procente*. În continuare vom prezenta două tipuri de reprezentări grafice: curba frecvenţelor cumulate şi curba procentelor cumulate.

Pentru exemplificare vom utiliza aceleaşi date din Tabelul 3.4. la care adăugăm trei coloane noi: limitele superioare ale claselor, frecvenţele cumulate şi procentajele cumulate, obţinând în final Tabelul 3.5.

Tabelul 3.5. Frecvenţele cumulate şi procentele cumulate aferente rezultatelor la testul matematic:

Cote brute x	Limita superioară a claselor	Frecvenţe f	Frecvenţe cumulate f_c	Procentaje cumulate %c
102-96	102,5	1	110	100,00
95-89	95,5	2	109	99,09
88-82	88,5	3	107	97,27
81-75	81,5	8	104	94,54
74-68	74,5	17	96	87,27
67-61	67,5	14	79	71,81
60-54	60,5	34	65	59,00
53-47	53,5	16	31	28,18
46-40	46,5	8	15	13,63
39-33	39,5	6	7	6,36
32-26	32,5	1	1	0.90
i=7		N=110		

Coloana frecvenţelor cumulate se obţine printr-o simplă adiţionare, începând de jos în sus. În prima clasă avem un singur caz, deci 1. Clasa a doua conţine şase cazuri la care adăugăm şi pe cele din clasele inferioare,

adică 1, rezultă pentru clasa a doua valoarea 7. În clasa a treia există 8 cazuri la care adăugăm numărul de cazuri inferioare, adică 7 şi obţinem 15 cazuri. În felul acesta se operează până la final, unde este obligatoriu să se ajungă la numărul total de subiecţi sau cazuri aflate în studiu.

Trebuie să subliniem faptul că, în cazul frecvenţelor cumulate nu este vorba de numărul de cazuri aferente unei clase sau alta, ci de numărul de cazuri aflate sub valoarea în cauză (limita superioară a clasei vizate).

La trasarea curbei frecvenţelor cumulate vom folosi aceleaşi axe rectangulare, doar că pe axa ordonatei vom trece limitele superioare ale claselor, iar pe abscisă marcăm frecvenţele cumulate, Figura 3.5.

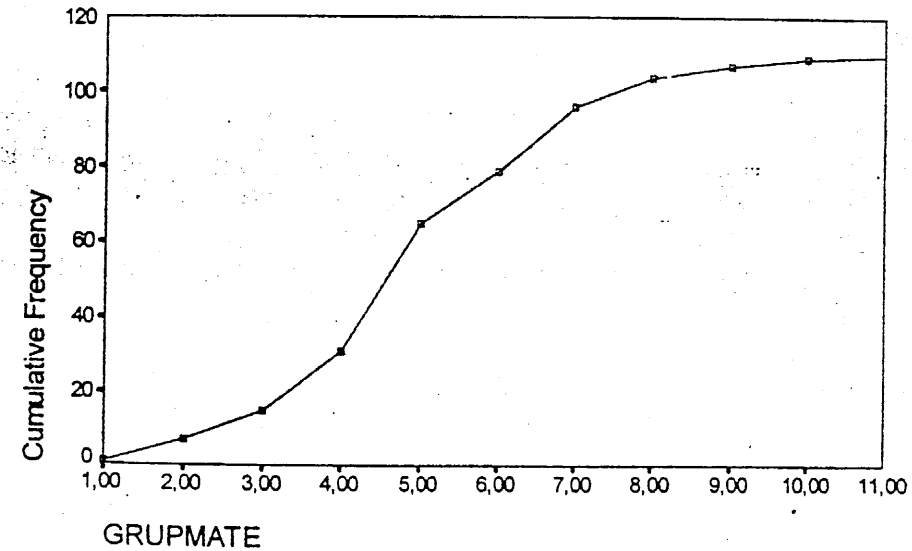


Fig.3.5. Curba frecvenţelor cumulate (sigmoida) la testul de cunoştinţe matematice

Exemplul 3.5. Să trasăm curba frecvenţelor cumulate a rezultatelor la testul de matematică existente în fişierul mate.sav (vezi Exemplul 3.3)

1. Din M **Graphs** se alege O **Line**...
2. În fereastra de **Line Charts** se alege modalitatea **Simple.**

3. Se apasă butonul **Define**; câmpul **Category Axis** se completează cu variabila GRUPMATE.
4. Pentru a trasa curba frecvențelor cumulate, în zona **Line Represents** se alege butonul **Cum. N of Cases**.
5. Se confirmă acțiunea prin **OK**.
6. Opțional, se poate salva graficul cu M **File**, O **Save As...**, dându-se numele frecvcum.cht.

3.2.5. Curba procentajelor cumulate

La trasarea acestei curbe se folosesc procentajele cumulate în locul frecvențelor cumulate pe axa abscisei, ordonata rămânând identică ca și la curba frecvențelor cumulate.

Procentajele cumulate (Tabelul 3.5.) nu sunt altceva decât frecvențe cumulate transformate în procente, utilizând regula de trei simplă.

Exemplu:

110----------100%
15----------x

$$x = \frac{15 \times 100}{110} = 13{,}63$$

110----------100%
96----------x

$$x = \frac{96 \times 100}{110} = 87{,}27$$

Utilizând datele înscrise în Tabelul 3.5. se trasează curba procentajelor cumulate sau ogiva, Figura 3.6.

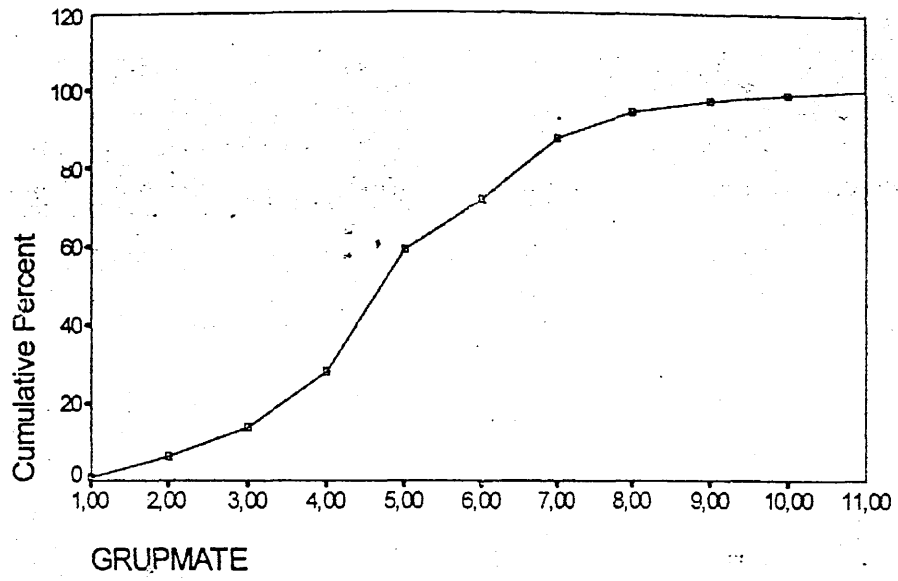


Fig.3.6. Curba procentajelor cumulate a rezultatelor la testul matematic.

Exemplul 3.6. Să trasăm curba procentajelor cumulate a rezultatelor la testul de matematică existente în fișierul mate.sav (vezi Exemplul 3.3).

1. Din M **Graphs** se alege O **Line...**
2. În fereastra de **Line Charts** se alege modalitatea **Simple.**
3. Se apasă butonul **Define**; câmpul **Category Axis** se completează cu variabila GRUPMATE.
4. Pentru a trasa curba frecvențelor cumulate, în zona **Line Represents** se alege butonul **Cum. % of Cases**.
5. Se confirmă acțiunea prin **OK**.
6. Opțional, se poate salva graficul cu M **File**, O **Save As...**, dându-se numele frecvcum.cht.

3.3. *Particularităţi ale distribuţiilor statistice*

În activitatea lor zilnică psihologii, sociologii şi pedagogii se întâlnesc cu o mare varietate de curbe reprezentând tot atâtea distribuţii de frecvenţe.

Practica statistică clasifică aceste distribuţii în funcţie de: centrul curbei, variaţia (împrăştierea datelor în jurul valorii centrale) şi aspectul general al curbelor în câteva forme particulare.

O formă frecvent întâlnită în practica ştiinţelor sociale şi în cea psihologică în special, este forma de ***clopot*** numită şi ***curba normală***. Această curbă se regăseşte în practică sub trei aspecte; Figura 3.7.

3.7. a) ***Distribuţie leptokurtică*** - când datele sunt grupate în jurul valorii centrale (mediei) fapt ce determină alungirea curbei pe verticală ;

3.7. c) ***Distribuţie platykurtică*** - se obţine atunci când datele din jurul mediei se răspândesc şi se împrăştie pe o zonă foarte extinsă, astfel că această curbă se extinde foarte mult pe orizontală ;

3.7 b) Între cele două extreme se află ***distribuţia mesokurtică***, care este curba normală specifică.

Fig. 3.7. Trei distribuţii sub formă de clopot:
a) - leptokurtică, b) - mesokurtică, c) - platykurtică.

Cele trei forme de curbe mai sus prezentate sunt ***simetrice***. Această simetrie presupune o repartiţie egală a valorilor frecvenţelor de o parte şi de alta a clasei centrale care conţine şi valoarea centrală.

Menţionăm că, există curbe simetrice, dar care nu au formă generală de clopot. Figura 3.8. prezintă trei astfel de exemple: a) - distribuţie

rectangulară, b) - distribuţie sub formă de U, c) - distribuţie bimodală (se întâlneşte când datele se grupează în jurul a două valori sau când un acelaşi fenomen se referă la două grupe distincte).

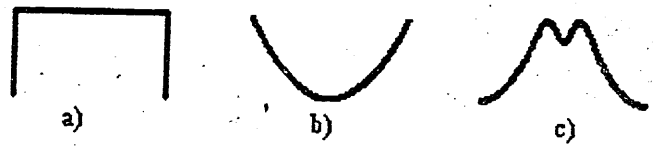


Fig. 3.8. Ilustrarea a trei curbe simetrice, dar care nu sunt normale.

De foarte multe ori, practicienii din domeniile psihologiei, sociologiei şi pedagogiei se întâlnesc cu un alt fel de curbă, respectiv cu distribuţii de frecvenţă ***asimetrice***. Dintre aceste tipuri de curbe două sunt tipice şi mai des întâlnite, Figura 3.9.

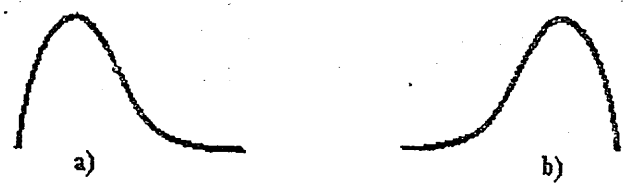


Fig.3.9. Două curbe asimetrice*: a) - orientare pozitivă, b) - orientare negativă (cotele cresc în valoare de la stânga la dreapta).

După cum se poate vedea în Figura 3.9., în aceste situaţii de distribuţii asimetrice este vorba de concentrarea frecvenţelor datelor spre una din cele două extremităţi.

3.4. *Consideraţii privind construcţia şi utilizarea graficelor*

Lecturarea oricărei cărţi de specialitate din domeniul psihologiei, sociologiei, pedagogiei şi practic din orice domeniu va demonstra necesitatea şi utilitatea prezentărilor intuitive ale datelor colectate

* după K.W.Elifson, R.P.Runyon şi A.Haber, pag. 71.

experimental, obținute în urma unor anchete sau testări. Statistica descriptivă pune la dispoziția celor interesați o serie de metode și tehnici de prezentare sub formă grafică a colecțiilor de date. Investigatorul va fi acela care va decide ce formă de prezentare este mai adecvată și se pretează cel mai bine fenomenului studiat.

Pentru a ușura această alegere, în cele ce urmează vom enunța câteva recomandări generale privind construcția și utilizarea graficelor:

- Histograma dă o reprezentare exactă a numărului de cazuri din fiecare clasă;
- În cazul histogramei, fiecare caz particular este reprezentat printr-o suprafață egală ca mărime;
- Histograma este utilă când este vorba doar despre o singură colecție de date sau o singură populație ;
- Poligonul frecvențelor are avantajul că descrie exact conturul distribuției (redă forma distribuției simetrică sau asimetrică);
- Datele individuale ale unei clase de intervale sunt concentrate într-un singur punct , respectiv valoarea centrală a clasei respective ;
- Poligonul frecvențelor permite compararea a două sau mai multe distribuții de frecvențe (sensibil egale ca și număr de subiecți, date brute);
- În cazul comparării a mai multor distribuții de frecvențe, dar inegale ca mărime, se recomandă utilizarea curbelor de procentaje cumulate ;

La construcția graficelor recomandăm să se țină seama de următoarele:

- Graficele ce ilustrează distribuții de frecvențe vor utiliza axele rectangulare. Pe axa absciselor (0x) se vor trece cotele brute sau scorurile iar pe axa ordonată (0y), frecvențele absolute ;
- Citirea graficelor se face de la stânga la dreapta. Pe scala orizontală (abscisă), valorile trebuie să crească de la stânga spre dreapta, iar pe scala verticală, de jos în sus. De altfel la intersecția axelor se află valoarea zero.
- Reperele scalelor se fixează arbitrar. Din punct de vedere estetic, se recomandă ca raportul înălțime - lățime să fie de 3/5.

- Cele două axe rectangulare vor purta câte o etichetă, denumirea a ceea ce reprezintă;
- Graficele vor fi numerotate și vor purta câte un titlu în partea de jos. În caz de nevoie se poate insera și câte o legendă cu explicații suplimentare.

3.5. Utilizarea tabelelor

O metodă foarte frecvent utilizată în prezentarea datelor statistice, rezultate dintr-o prelucrare primară o constituie ***tabelarea.*** *Tabelele trebuiesc astfel concepute încât să permită persoanei care le citește să surprindă rapid și ușor informația transmisă.*

Tabelele se introduc în manuscris pe baza unui număr care se trece în partea stângă sus, iar în continuare va conține un titlu sintetic al conținutului.

ATENȚIE la modul în care sunt trecute datele într-un tabel. Nu orice fel de date fac tabelul inteligibil. Construcția tabelului este determinată de scopul cercetării și realizată astfel încât aspectele care urmează să fie relevante.

Redăm foarte pe scurt câteva reguli generale de alcătuire a tabelelor, care trebuie luate în considerare în mod obligatoriu:

- Tabelele trebuie să conțină în partea stângă sus un număr de identificare și un titlu de prezentare, clar, sintetic, precis.
- În caz de nevoie explicațiile suplimentare se trec imediat sub titlu sau la subsolul tabelului.
- Liniile și coloanele tabelului trebuie așezate în ordine logică pentru a facilita efectuarea de comparații.
- Pentru a se evita comiterea de confuzii și erori în urmărirea datelor se recomandă gruparea datelor câte 5 sau 10 cu un spațiu liber între ele.
- Dacă cifrele de prezentat în tabele sunt foarte lungi, se recomandă gruparea lor. Exemplu: cifra 29673584 se va trece 29 673 584.
- O ultimă recomandare este de ordin estetic. Prima și ultima linie a tabelului vor fi duble sau îngroșate. Tabelele nu se vor înghesui, dar nu se vor lăsa nici spații excesive.

se face apel la compararea datelor obținute de la două sau mai multe grupe inegale ca număr. În aceste situații transpunerea în tabel a frecvențelor brute nu constituie o bază reală de comparație. Aceste situații cer ca datele brute să fie înregistrate ***procentual***.

De exemplu: Un sociolog efectuând o cercetare în domeniul reclamei, a experimentat trei moduri diferite de prezentare a aceluiași tip de șosete pentru copii. Cele trei moduri de prezentare sunt:

- modul I - în perioada 01-31.I.1996 a expus șosetele într-un coș de nuiele.
- modul II - în perioada 01-31.III.1996 a expus șosetele în pungi de plastic câte trei în pungă
- modul III - în perioada 01-31.V.1996, șosetele au fost expuse în cutii cu un colorit deosebit (o pereche / cutie).

Toți vizitatorii magazinului în perioadele respective au fost intervievați despre diversitatea produselor din unitate. Investigația s-a făcut pe baza unui chestionar din care 15 itemi se referau exclusiv la produsul în cauză (șosete pentru copii).

Rezultatele finale sunt redate în Tabelul 3.6.

Tabelul 3.6. Rezultatele investigației vizând cele trei moduri de prezentare a șosetelor pentru copii.

MODUL DE PREZENTARE	NUMĂR DE SUBIECȚI	PREFERINȚE EXPRIMATE
I	70	560
II	60	540
III	80	600

Simpla însumare a răspunsurilor (în cazul nostru preferințe) poate crea o imagine falsă asupra rezultatelor. Din Tabelul 3.6. rezultă că modul III este cel mai eficient, urmat de modul I, apoi de modul II. Eroarea este introdusă de utilizarea datelor brute, absolute, raportate la efective inegale deci necomparabile în actualul context.

formă procentuală, utilizând indicele numit ***procent de exactitate (PE)***.

$$PE = \frac{R_c \times 100}{R_{max}} \qquad (3.1)$$

PE - procent de exactitate

R_c - răspunsuri corecte

R_{max} - număr maxim de răspunsuri corecte

În exemplul nostru, chestionarul conține 15 itemi. Răspunsul maxim va fi obținut din produsul 15 cu efectivul clasei în cauză.

Astfel se va obține:

Pentru modul I:

$$PE_I = \frac{560 \times 100}{70 \times 15} = 53{,}3$$

Pentru modul II:

$$PE_{II} = \frac{540 \times 100}{60 \times 15} = 60$$

Pentru modul III:

$$PE_{III} = \frac{600 \times 100}{80 \times 15} = 50$$

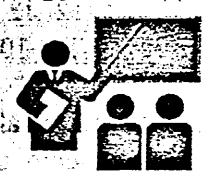
Utilizând datele obținute vom putea întocmi un nou Tabel 3.7., care de această dată conține datele în formă comparabilă și concluzia reală privind preferințele modurilor de prezentare.

Tabelul 3.7. Răspunsurile referitoare la preferințele privind modul de expunere a șosetelor de copii.

Modul de prezentare	Număr de subiecti	Procent de exactitate
I	560	53,3
II	540	60,0
III	600	50,0

Concluzie: cel mai eficient mod de prezentare este cel de la modul II urmat în ordine de modurile III şi I. Menţionăm că în caz de nevoie pe lângă procentul de exactitate pot fi trecute în paranteze şi cifrele absolute.

Capitolul 4. DETERMINAREA TENDINŢEI CENTRALE



- **media**
- **mediana**
- **modul**

În capitolul anterior au fost prezentate o serie de tehnici prin care o colecţie de date brute, aleatoare pot fi aranjate în aşa fel încât să devină inteligibile şi să ne furnizeze o serie de informaţii destul de utile, constituind punctul de plecare în analiza statistică.

Pe lângă posibilitatea ca aceste date brute să fie prezentate sub formă de tabele şi/sau reprezentări grafice, este nevoie şi de o exprimare ***sintetică*** şi ***cuantificată*** a unui ansamblu sau altul de date.

De exemplu, foarte des în pedagogie se utilizează expresia: "Eleva X este mai bună decât eleva Z, dar mai slabă decât eleva Y".

Care este baza acestei afirmaţii, la ce se referă cel care o face? În fiecare clasă găsim aceleaşi note şcolare de la 4 la 10. Atunci când facem astfel de aprecieri ne raportăm la *valoarea medie* a performanţelor care caracterizează fiecare colectivitate în parte, în cazul de faţă media generală a clasei.

În cercetările psihologice, sociologice şi pedagogice în permanenţă cercetătorul este obligat să determine valoarea centrală a ansamblului de date obţinut experimental. În practică există mai mulţi indici statistici prin care se poate exprima valoarea centrală, aceştia sunt : media aritmetică (pe scurt media), mediana şi modul.

4.1. Media aritmetică

În statistică şi nu numai, ***media aritmetică*** sau pe scurt ***media*** se notează cu litera "***m***". Această medie este un descriptor al unui ansamblu de date aparţinătoare unui eşantion extras dintr-o populaţie.

În practică se întâlnesc trei situaţii specifice, caracteristice, care i-au determinat pe statisticieni să dezvolte trei tehnici mai des utilizate.

4.1.1. Calculul mediei pe date negrupate.

Să luăm un exemplu: Un psiholog în timpul determinării timpului de reacție la un subiect a înregistrat următoarele valori: 11, 13, 12, 14, 12, 12, 13, 13, 15, 11, 12, 14, 11, 12, 13, 13, 14, 15, 10, 17, 10, 12, 16, 18, 10 sutimi de secundă.

Se ridică întrebarea: care este valoarea timpului de reacție real al subiectului examinat? Răspunsul este: trebuie căutată valoarea medie a ansamblului de date.

În mod practic, media aritmetică, media este: *suma unui set de măsurători împărțită la numărul total al acestora.*

Formula este:

$$m = \frac{\sum x}{N} \qquad (4.1.)$$

m = media

Σ = "sumă de"

x = valorile individuale

Revenind la exemplul nostru:

$$m = \frac{\sum x}{N} = \frac{11+13+12+14+12+12+13+13+15etc}{25} = \frac{325}{25} = 13$$

Răspunsul final este: timpul de reacție al subiectului este egal cu 13 sutimi de secundă.

4.1.2. Calculul mediei unei distribuții de frecvență

Media mai sus descrisă este eficientă și operatorie doar în situațiile când numărul de date este mic. Frecvent cercetătorul este confruntat cu situații în care numărul datelor este de ordinul zecilor. Astfel de cazuri reclamă în primul rând o distribuție de frecvență, iar în al doilea rând o formulă adecvată acestor situații.

$$m = \frac{\sum fx}{N} \qquad (4.2)$$

f = frecvența unei valori

x = valoarea individuală

fx = produsul dintre frecvență și valoarea individuală aferentă

Exemplu: Aplicându-se un test de cunoștințe la o clasă de elevi s-au obținut următoarele rezultate:

3	4	5	6	7	7	7
5	3	6	7	8	8	8
6	6	4	8	7	8	7
7	6	6	9	8	7	9
6	5	7	7	8	8	10
9	10	9	3	10	5	10
8	4	5	5	7	4	6
9	6	4	10	9	7	9
10	9	8	4	7	6	5
4	6	6	3	5	7	7

Primul pas este construirea unei distribuții de frecvență:

Tabelul 4.1. Distribuția de frecvență a performanțelor la testul de cunoștințe.

x	f	f_x
10	6	60
9	8	72
8	10	80
7	15	105
6	12	72
5	8	40
4	7	28
3	4	12
i=1	N=70	Σf_x=469

Utilizând formula 4.2 și înlocuind cu datele Tabelului 4.1 vom obține:

$$m = \frac{\sum fx}{N} = \frac{469}{70} = 6,7$$

Răspuns: media performanţelor obţinute de elevii clasei este 6,7.

Exemplul de mai sus, când $i = 1$, se întâlneşte mai rar. În mod obişnuit i este diferit de 1, cum este în exemplul din Tabelul 4.2.

Tabelul 4.2. Distribuţia de frecvenţă a performanţelor la un test de spirit de observaţie.

x	x_k	f	f_{xk}
26-24	25	1	25
23-21	22	3	66
20-18	19	4	76
17-15	16	10	160
14-12	13	16	208
11-9	10	9	90
8-6	7	5	35
5-3	4	3	12
i=3		N=51	$\sum fxk = 672$

Formula pe care o utilizăm în astfel de situaţie este uşor modificată şi arată astfel:

$$m = \frac{\sum fxk}{N} \qquad (4.3)$$

Înlocuind cu datele exemplului nostru vom obţine:

$$m = \frac{\sum fxk}{N} = \frac{672}{51} = 13{,}176$$

Notăm faptul că în situaţiile în care se obţin zecimale se recomandă şi se obişnuieşte rotunjirea la o singură zecimală. Rezultatul obţinut 13,176 se va scrie m=13,2.

Răspunsul: media performanţelor obţinute la testul de observaţie este: m=13,2

4.1.3. Determinarea mediei utilizând valori codificate

Metodele de determinare, de calculare a mediei aritmetice pe care le-am descris până acum , presupun utilizarea unei tehnici de lucru foarte laborioase, mai ales când numărul de cazuri, de cote brute este de ordinul sutelor. Pentru astfel de situaţii s-a elaborat o metodă mai scurtă, bazată pe unele proprietăţi algebrice, implicând utilizarea de valori codate.

Exemplu: În cadrul unei baterii de probe psihologice vizând nivelul intelectual general s-a aflat şi o probă de intuiţie tehnică care a fost aplicat unui eşantion de 220 de subiecţi. Rezultatele obţinute sunt:

56	58	54	72	56	67	99	75	68	75	71
49	58	58	60	59	61	90	40	51	40	58
60	67	81	34	44	68	54	47	54	48	59
56	59	62	57	85	46	48	71	47	63	62
65	82	75	71	54	68	58	61	64	58	34
67	54	70	45	58	60	50	72	58	64	48
55	48	74	48	48	39	57	73	51	58	71
54	62	50	51	61	58	42	56	37	37	61
36	26	56	78	46	69	77	94	74	52	56
68	59	54	75	55	46	70	72	87	77	52

67	99	56	68	52	58	77	64	75	61	71
61	90	59	36	56	67	37	47	62	71	57
68	54	44	54	61	58	52	54	81	40	34
46	48	85	55	71	59	58	51	58	47	60
68	58	54	67	48	82	64	68	54	75	72
60	50	58	65	34	48	63	58	70	72	45
39	57	48	56	62	54	58	51	74	73	48
58	42	61	60	59	26	40	37	50	56	51
69	77	46	49	58	62	48	47	56	94	78
48	70	55	56	71	59	75	87	54	72	75

Ne interesează să determinăm media eşantionului examinat. Procedeul practic de calcul este următorul:

Pasul 1. Făcând uz de cunoştinţele dobândite în capitolele anterioare, construim un tabel care să conţină distribuţia de frecvenţă a datelor de mai sus, Tabelul 4.3.

x	x_k	f	u	$f*u$
99-95	97	2	8	16
94-90	92	4	7	28
89-85	87	4	6	24
84-80	82	4	5	20
79-75	77	15	4	60
74-70	72	22	3	66
69-65	67	18	2	36
64-60	62	26	1	26 276
59-55	57	48	0	-170
54-50	52	28	-1	-28
49-45	47	27	-2	-54
44-40	42	8	-3	-24
39-35	37	8	-4	-32
34-30	32	4	-5	-20
29-25	27	2	-6	-12
i=5		N=220		Σf*u=106

Pasul 2. Acest pas constă în efectuarea codării, respectiv construirea coloanei "***u***". În primul rând se alege arbitrar un punct de referință care obișnuit se notează cu m_0 și este valoarea centrală a clasei cu frecvența cea mai mare. În cazul nostru, clasa cu frecvența maximă este 59-55 cu valoarea centrală $x_k = 57$. Acest punct (m_0) se alege în funcție de unde se estimează că se află media, știind că aceasta se află în punctul cel mai înalt al curbei (poligonului frecvenței) respectiv în intervalul de clasă cu frecvența maximă.
În dreptul acestei valori ($x_k = 57$) se trece valoarea zero în coloana "u".

Următoarea operație este numerotarea liniilor de clasă (x) în jos și în sus față de punctul zero (1,2,3,4....etc.) până la epuizarea acestora. În continuare se va acorda semnul plus acelor linii unde valoarea absolută a variabilei x crește față de acest punct zero (în cazul nostru în sus). Semnul

minus se trece în direcția descreșterii valorilor brute față de același punct zero (în cazul nostru în jos).

Astfel s-a obținut coloana "u" a Tabelului 4.3.

Notă: Această operație de codare se efectuează în mod mecanic în orice fel de situație, dar are la bază o explicație matematică, formula de calcul fiind:

$$u = \frac{x_k - m_0}{i} \qquad (4.4)$$

De exemplu: pentru clasa 79 - 75 cu $x_k = 77$ se va obține:

$$u = \frac{77 - 57}{5} = 4$$

sau pentru clasa 34 - 30 cu $x_k = 32$ se va obține:

$$u = \frac{32 - 57}{5} = -5$$

Pasul 3. Acesta constă în întocmirea coloanei fu, adică efectuarea produsului pentru fiecare interval de clasă între frecvență (f) și valoarea codată (u). Apoi se calculează suma algebrică a valorilor coloanei, obținându-se suma coloanei, care în exemplul nostru este: $\Sigma fu = 106$.

Pasul 4. Aplicăm formula de calcul a mediei care este:

$$m = m_0 + i\frac{\sum fu}{N} \qquad (4.5)$$

Făcând înlocuirile în formulă cu datele trecute în Tabelul 4.3 vom obține:

$$m = 57 + 5\frac{106}{220} = 59{,}409$$

Aplicând rotunjirea vom obține: m = 59,4.

Deci, media performanţelor la testul de intuiţie tehnică a eşantionului de 22 subiecţi este: m = 59,4.

Exemplul 4.1. Calculul mediei performanţelor la testul de intuiţie tehnică a eşantionului de 220 subiecţi.

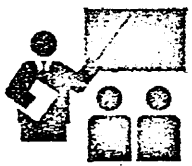
1. Se crează variabila PERFORM, inserând toate rezultatele celor 220 de subiecţi
2. Din M **Statistics,** submeniul **Summarize** se alege **Descriptives....** În lista de variabile **Variables** se trece variabila PERFORM.
3. Se apasă butonul **Options** şi se bifează **Mean** pentru calculul mediei.
4. Se salvează fişierul cu numele intuitie.sav.

***Observaţie*:** Calculul mediei se poate face şi cu ajutorul M **Statistics**, submeniul **Summarize,** O **Frecquencies**. Tot ca mai înainte, se trece variabila PERFORM în lista de variabile, se apasă butonul **Statistics...,** iar aici se va bifa pentru calculul mediei, **Mean** din zona **Central Tendency**.

Se va obţine:

Number of valid observations (listwise) = 220,00

Variable	Valid Mean	N	Label
PERFORM	59,46	220	



4.1.4. Combinarea mediilor provenite de la eşantioane diferite.

De foarte multe ori în practică apare situaţia când suntem nevoiţi să aglutinăm două sau mai multe grupuri independente şi în consecinţă trebuie să obţinem o medie comună.

Pentru acest lucru se recomandă utilizarea următoarei formule:

$$M = \frac{n_1 m_1 + n_2 m_2 + ... + n_k m_k}{N} \qquad (4.6.)$$

M - media generală a grupelor reunite;
$m_1, m_2....m_k$ - media grupelor independente;
$n_1, n_2, ...n_k$ - numărul subiecţilor grupelor;
N - numărul total de subiecţi ($N = n_1+n_2+.....n_k$).

Exemplu: În urma aplicării unui test de aptitudini generale, Wonderlic la trei grupe independente de subiecţi sau obţinut rezultatele: $n_1 = 30$, $m_1 = 21,5$; $n_2 = 45$, $m = 20,0$; $n_3 = 18$, $m_3 = 20,8$. Dorim să contopim cele trei grupe în unul singur şi ca atare ne interesează un singur indicator al mediei. Pentru acest lucru vom utiliza Formula: 4.6:

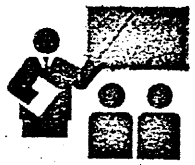
$$M = \frac{(30 \times 21,5) + (45 \times 20,0) + (18 \times 20,8)}{30+45+18} = 20,6$$

Media comună a celor trei grupe reunite este: M = 20,6.

Formula de contopire a mediilor provenite de la grupe independente se poate extinde la k grupe în funcţie de caz şi necesităţi.

Exemplul 4.2 Să calculăm media comună a celor trei eşantioane considerate în exemplul anterior.

- *Sugestie:* Se lansează aplicaţia Calculator şi se utilizează ca şi calculatorul de buzunar



4.1.5. Media procentajelor.

Apar şi cazuri în care suntem obligaţi să calculăm media a două sa mai multe procentaje. Dacă, numărul de date este sensibil apropiat se poat utiliza una din formulele de calcul al mediei.

În cazul în care numărul de date este foarte diferit, este nevoie d utilizarea, respectiv calcularea unei medii ponderate. Acest lucru est obligatoriu pentru că, astfel este anihilată eroarea introdusă de lucrul c eşantioane şi procentaje inegale.

Exemplu: Suntem interesaţi să aflăm procentajul mediu de respinşi l examenul de admitere la psihologie a candidaţilor proveniţi din diferit localităţi. Situaţia sintetică de start este prezentată în Tabelul 4.4.

Tabelul 4.4.Situaţia elevilor respinşi la examenul de admitere la psihologie

Localităţi	Nr. elevi prezentaţi la admitere N_i	Nr. elevi respinşi $N_iP_i/100$	% respinşi P_i
A	243	55	22,6
B	63	7	11,1
C	196	43	21,9
D	61	2	3,3
E	125	34	27,2
Total	$\Sigma N_i = 688$	$\Sigma N_iP_i / 100 = 141$	$\Sigma P_i = 86,1$

Pentru calcularea mediei ponderate a procentajelor vom utiliza formula:

$$Mp\% = \frac{\sum N_i P_i}{\sum Ni} \quad (4.7)$$

unde:

N_i - numărul de cazuri per eşantion;

P_i - procentul fiecărui eşantion

$N_iP_i/100$ - produsul fiecărui procentaj cu fiecare N parţial împărţit la 100.

Operând înlocuirile în formulă cu datele noastre concrete din Tabelul 4.4 vom obţine:

$$Mp\% = \frac{14.100}{688} = 20,5$$

Răspunsul care se dă la întrebarea formulată în exemplul nostru este că media procentajului de respinşi este de : $M_{p\%} = 20,5$.

În cazul în care se lucrează cu proporţii în loc de procentaje se va aplica formula:

$$Mpp = \frac{\sum N_i P_i / 100}{\sum N_i} \quad (4.8)$$

După cum se poate lesne constata apare o simplă înlocuire a procentajelor cu proporţiile corespunzătoare.

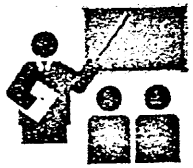
Exemplul 4.4. Să calculăm procentajul mediu al elevilor respinşi la examenul de admitere la psihologie, cu datel din exemplul de mai sus.

1. Se crează trei variabile NRPREZ, NRRESP, PROCRESP
2. Din M **Data** alegem O **Weight Cases...**; vom selecta butonul **Weight Cases by**, iar câmpul **Frecquency variable** îl vom completa cu variabila NRPREZ. În acest mod am realizat ceea ce se cheamă o ***ponderare***, toate variabilele fişierului fiind ponderate corespunzător prin valorile variabilei de ponderare, în cazul nostru NRPREZ. În urma ponderării, va apare în colţul dreapta jos un mesaj, **Weight on**, ceea ce ne atenţionează asupra faptului că este activă ponderarea.
3. Vom calcula media ponderată a procentajelor cu ajutorul M **Statistics**, submeniul **Summarize**, O **Descriptive**, alegând variabila care ne interesează, adică PROCRESP.
4. Se salvează fişierul cu numele elevresp.sav

Se va obţine:

Number of valid observations (listwise) = 688,00

Variable	Valid Mean	N	Label
PROCRESP	20,47	688	



4.2. Mediana

Un indice foarte des utilizat ca şi indicator al tendinţei centrale este ***mediana (Md)***. Ca şi definiţie, *mediana este valoarea care împarte un şir de date ordonate în două grupe egale ca şi număr*. Practic, este punctul deasupra şi dedesubtul căruia se situează 50% din totalul valorilor.

Atenţie:

- *prima operaţie când intenţionăm să determinăm mediana este ordonarea întregii colecţii de valori (crescător sau descrescător).*
- *a doua remarcă se referă la faptul că, mediana se determină în situaţiile în care distribuţia de frecvenţă este asimetrică.*

4.2.1. Mediana datelor negrupate.

În astfel de cazuri se disting două faze:
a) identificarea *locului* medianei
b) determinarea *valorii* medianei.

Cazul datelor negrupate prezintă trei situaţii caracteristice

- Cazul în care şirul ordonat de date este **impar**:

Exemplu: Avem valorile: 17, 18, 22, 25, 26, 27, 31, 34, 39.
În primul rând determinăm locul medianei cu formula:

$$M_d = \frac{N+1}{2} \quad (4.9)$$

În cazul nostru locul medianei este:

$$M_d = \frac{9+1}{2} = 5$$

Deci, a cincea poziţie din şirul ordonat de date este mediana. În şirul nostru poziţia a 5-a este 26. Concluzia: mediana şirului nostru de date este: $M_d = 26$.

- Cazul în care şirul de date este **par**, constituie a doua situaţie specifică în determinarea medianei.

Exemplu: Într-o testare psihologică s-au obţinut valorile: 27,28,32,35,36,37,41,44,49,50. Ca şi în situaţia anterioară determinăm în primul rând locul medianei. În cazul nostru acest loc este: 5,5. Cu alte cuvinte valoarea medianei este pe locul 5,5, adică ea se află între a 5-a şi a 6-a valoare. Media acestor două valori va fi mediana căutată în această situaţie, care este: $M_d = (36 + 37) / 2 = 36,5$.

- A treia posibilitate întâlnită în situaţii când datele brute nu sunt grupate este atunci când anumite valori ale şirului **se repetă**.

Exemplu: 57,58,59,64,64,64,67,67,69. Şi în această situaţie se procedează identic ca şi în cele anterioare. De această dată locul medianei este 5. Pe locul cinci din şirul de date ordonat se află mediana căutată, care este: M_d= 64.

4.2.2. Mediana datelor grupate

În mod obişnuit testările sau examinările psihologice, anchetele sociale se efectuează pe eşantioane numeroase, necesitând gruparea datelor, iar determinarea medianei este diferită faţă de situaţiile menţionate anterior

Exemplu: În cadrul unei cercetări de psihologie diferențială, M.P.Manson aplicat un chestionar de alcoolism/nonalcoolism unui eşantion reprezentat pentru populația studențească din Universitatea "Southern", Californi S.U.A. Rezultatele obținute sunt:

13	10	1	2	1	0	6	7	6	2	9
0	10	1	12	13	19	1	30	14	22	1
14	19	25	16	19	10	25	20	0	1	3
2	2	3	1	5	2	3	5	6	7	1
1	9	10	9	13	17	10	0	11	14	12
9	12	13	1	6	1	7	13	7	13	1
15	16	1	14	10	9	12	19	7	2	9
19	6	20	21	19	11	2	17	1	22	20
25	0	25	27	15	22	9	10	9	19	24
9	27	6	6	6	27	19	15	11	15	2
28	9	7	41	58	12	59	0	9	25	21
25	13	48	49	22	15	19	1	17	5	57
1	0	2	44	9	50	27	2	20	9	44
27	24	22	6	7	12	10	11	9	15	1
10	15	6	16	17	9	19	5	1	1	2
1	12	0	14	11	15	16	17	11	19	10
20	19	13	22	2	17	24	22	20	20	1
17	36	28	12	9	3	25	19	13	20	20
2	20	44	19	17	9	11	19	14	20	10

După cum am explicat şi motivat în capitolele anterioare, primul p îl constituie întocmirea distribuției de frecvență a datelor brute, obținându- Tabelul 4.5.

Tabelul 4.5. Distribuția de frecvență a rezultatelor chestionarului alcoolism/nonalcoolism"

Cote brute x	Limita inferioară	Valoarea centrală x_k	Frecvența f	Frecvența cumulată f_c
0 - 4	-0,5	2	45	210
5 - 9	4,5	7	39	165
10 - 14	9,5	12	41	126
15 - 19	14,5	17	35	85
20 - 24	19,5	22	24	50
25 - 29	24,5	27	14	26
30 - 34	29,5	32	1	12
35 - 39	34,5	37	1	11
40 - 44	39,5	42	4	10
45 - 49	44,5	47	2	6
50 - 54	49,5	52	1	4
55 - 59	54,5	57	3	3
i = 5			N = 210	

În astfel de cazuri, când datele sunt grupate, mediana este determinată în mod direct prin utilizarea formulei:

$$M_d = L + i\frac{\frac{N}{2} - F}{f_m} \qquad (4.10)$$

unde: L - limita inferioară a clasei unde este situată mediana;
F - numărul de valori inferioare lui L;
f_m - frecvența clasei care include valoarea medianei.

Pentru exemplul nostru putem identifica următoarele:

1. $N/2$=210/2 este egal cu = 105, locul medianei care este în intervalul de clasă 10 - 14;
2. Limita inferioară a clasei care conține mediana este: L = 9,5;
3. F - numărul de valori inferioare lui L este egală cu 85;
4. Din Tabelul 4.5 identificăm indicatorul f_m = 41.

Având identificate valorile cerute de formula de calcul putem trece la înlocuirile aferente în urma căruia vom obține:

$$M_d = 9{,}5 + 5\frac{105-85}{41} = 11{,}90$$

Concluzia: mediana colecției de date, a rezultatelor obținute în urma aplicării chestionarului de alcoolism-nonalcoolism este: $M_d = 11{,}90$.

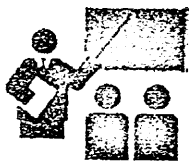
În funcție de situația concretă, experimentatorul va putea utiliza metoda cea mai adecvată de determinare a medianei ca și indicator central al unei distribuții de frecvență (ordonată sau grupată).

Exemplul 4.5. Ne propunem să calculăm mediana pentru rezultatele obținute în urma chestionarului de alcoolism-nonalcoolism.

1. Se crează variabila VAR cu datele din chestionar..
2. Din M **Statistics** submeniul **Summarize** se alege O **Frecquencies**. Se introduce variabila VAR în câmpul **Variables**.
3. Se apasă butonul **Statistics...**, iar pentru calculul medianei se va bifa opțiunea **Median** din zona **Central Tendency**.
4. Se salvează fișierul cu numele alcol.sav.

Se va obține:

```
Median          12,000
```



4.3 Modul

O măsură mai rar utilizată ca și indicator al tendinței centrale este ***modul***, prescurtat Mo. Ca și definiție, *modul este valoarea a cărei frecvență este cea mai mare în colecția de date.*

În principal, modul se determină prin simpla inspectare vizuală a colecției de date. În cazul datelor grupate, modul este valoarea centrală (x_k) a intervalului de clasă cu frecvența cea mai mare.

Să luăm două **exemple**:

Avem următoarea serie de date: 12, 13, 13, 13, 15, 17, 17, 17, 17, 17, 18, 18, 18, 19, 19, 21, 22.

În această situație modul este Mo = 17., deoarece este valoarea cu cea mai mare frecvență, adică 5 (se repetă cel mai des).

Dacă datele sunt grupate, ca și în cazul Tabelului 4.3, modul este valoarea centrală a intervalului de clasă cu frecvența cea mai mare, cu alte cuvinte: Mo=57.

În practică, se întâlnesc câteva situații concrete în care **NU** se poate identifica modulul. Acestea sunt:

a) în seriile în care frecvența este egală cu 1. De exemplu: 27, 29, 30, 35, 38, 39, 40.

b) de asemenea, nu putem identifica modulul dacă fiecare valoare are aceeași frecvență ca și în seria:122, 122, 125, 125, 131, 131, 132,132

c) o situație aparte este atunci când două valori alăturate au aceeași frecvență . În astfel de situații, modul este valoarea intermediară a acestora. De pildă:

92, 92, 93, 95, 96, 96, 96, 97, 97, 97, 98, 98, 99, 99.

În această situație , valorile 96 și 97 au aceeași frecvență 3. Modul șirului va fi valoarea intermediară, adică: Mo = (96+97)/2=96,5

d) în cazul în care șirul de valori conține două date cu aceeași frecvență, dar care se află la distanță una de alta, vom avea două module. De exemplu:

23, 23, 23, 24, 24, 24, 24, 25, 25, 25, 26

26, 27, 28, 28, 28, 28, 28, 29, 29, 30, 31

După cum se vede, valorile 24 și 28 au frecvența 4. În consecință, vom avea două moduri: Mo_1=24 și Mo_2=28. Acest lucru se întâmplă atunci când avem de-a face cu distribuții bimodale.

Notă: Modul este un indice aproximativ, nu se recomandă utilizarea lui în prelucrări statistice, are doar o valoare orientativă.

4.4. Media - mediana - modul - Considerații finale -

După cum s-a arătat și la începutul acestui capitol, media - mediana - modul sunt indicatori, ***măsuri ale tendinței centrale*** într-o distribuție de frecvențe. Pentru rigurozitate în statistică, se cere obligatoriu nominalizarea indicatorului concret și nu utilizarea noțiunii de tendință centrală, care poate genera neînțelegeri și discuții inutile.

Exemplu: Situația salariului pe un an într-o companie oarecare este prezentată în tabelul 4.6. (după K. W. Elifson și ceilalți).

Tabelul 4.6. Salariul anual în compania XYZ (în dolari)

Salar (dolari/an)	Tendința centrală
30.000	
24.000	
14.000	
10.000	← Media
9.000	
9.000	
9.000	
7.000	← Mediana
6.000	
6.000	
6.000	
5.000	
5.000	← Modul
5.000	
5.000	

Muncitorii companiei declară grevă în vederea măririi salariului, ei susținând că salariul mediu este de 5.000 de dolari. Managerul nu este de acord, motivând că după calculele lui, salariul mediu este de 10.000 de dolari. Un observator extern notând aceleași date, remarcă: "Amândoi greșiți. Salariul mediu este de 7.000 de dolari".

Considerăm că exemplul de mai sus este suficient de elocvent pentru sublinia necesitatea precizării indicatorului central utilizat. Mai mult, rezultă cu mare claritate faptul că media, mediana și modul au semnificații diferite.

- ◊ ***Media*** - este măsura cea mai adecvată a centrului unei serii de variație. Ea surprinde integral ansamblul de date, cote brute. În același timp, permite utilizarea unui larg evantai de prelucrări algebrice. Media se determină în cazul distribuțiilor simetrice.
- ◊ ***Mediana*** - centrul, mijlocul unui ansamblu de date ordonate. Calculul medianei se bazează pe proprietățile ordinale ale datelor. Mediana este preferată în distribuțiile asimetrice, deoarece face abstracție de valorile extreme.
- ◊ ***Modul*** - valoarea cu frecvența maximă pe abscisă, corespunde cu punctul cel mai ridicat al curbei frecvență. Este rar utilizat, doar în situațiile de distribuții bi și multimodale, utilizări de variabile calitative, fragmente de curbe etc.

Cei trei indicatori, media, mediana și modul ne sugerează dacă o distribuție este sau nu simetrică, iar în caz de asimetrie, ne indică și direcția pozitivă sau negativă. Atunci când cei trei indicatori au aceeași valoare, avem de-a face cu o distribuție simetrică, Figura 4.1.c. În cazul în care media este mai mare decât mediana și amândouă mai mari decât modul, avem o distribuție asimetrică orientată pozitiv, Fig. 4.1.a. Dacă media este mai mică decât mediana și ambele decât modul, distribuția este de tip asimetrică negativă, Fig. 4.1.b.

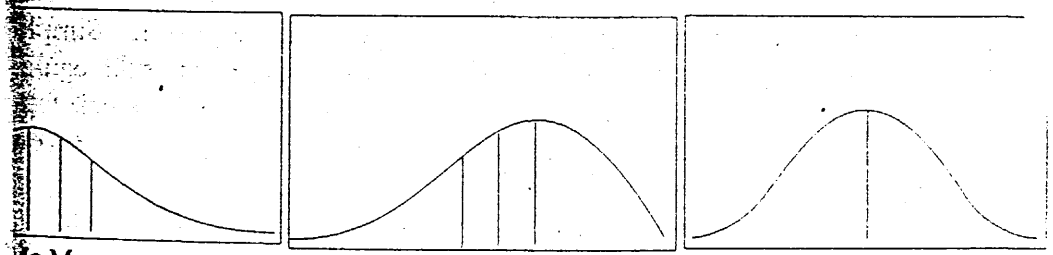
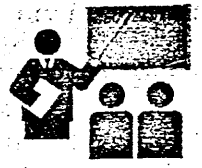


Fig. 4.1. Relația dintre medie, mediană, mod și forma distribuției de frecvențe.

Capitolul 5 MĂSURI ALE DISPERSIEI



- **rangul**
- **abaterea medie**
- **abaterea standard**
- **grad de libertate**
- **variabilitate**

Un scor oarecare, prin sine însuşi nu are nici o semnificaţie. Scor poate deveni semnificativ doar când este comparat cu un alt scor. În cazul care cunoaştem media unei distribuţii de frecvenţe, se poate determina da un scor oarecare este mai mare sau mai mic decât media variabilei în cauz Dar, cât de mare sau mic este? Este limpede că, având doar indicator tendinţei centrale, vom putea obţine doar informaţii limitate în acest sen Pentru a descrie o distribuţie mult mai complet sau pentru a interpreta mi mai precis un scor oarecare este nevoie şi de o altă informaţie care es concentrată în ***dispersia*** scorurilor în jurul mediei.

De exemplu: Aplicându-se un test de relaţii spaţiale la două grupe subiecţi, s-au obţinut următoarele rezultate:

Grupul A: 22, 28, 30, 32, 35, 35, 38, 40, 42, 48

Grupul B: 20, 22, 28, 30, 32, 32, 38, 45, 45, 58

Ambele grupe au obţinut o performanţă medie de m = 35. Simp inspecţie vizuală a datelor ne relevă că cele două grupe nu sunt egal Grupul A este mai omogen, datele sunt relativ grupate în jurul mediei, fa de grupul B unde datele sunt împrăştiate, dispersate pe o plajă mai mare, alte cuvinte, grupul B este mai neomogen.

Să luăm un alt exemplu, ilustrat în Figura 5.1., după K.W.Elifson.

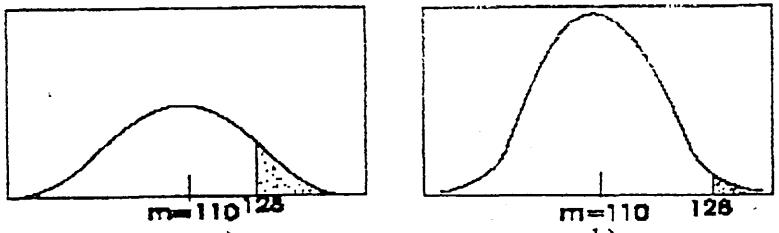


Fig. 5.1. Două curbe de frecvenţă cu aceeaşi valoare a mediei, dar cu dispersie diferită.

În ambele distribuţii, media este aceeaşi: m = 110, dar există o semnificativă diferenţă în ceea ce priveşte interpretarea cotei 128. În Figura 5.1.a), scorul 128 poate fi considerat destul de apropiat de medie, având o mărime moderată ca şi valoare a variabilei. Acelaşi scor, 128, în Figura 5.1.b) este considerat a fi în vârful topului; cota este foarte ridicată.

Din cele arătate până acum rezultă că, pentru a interpreta un scor individual oarecare este nevoie pe lângă medie şi de un indicator care să exprime gradul de dispersie, împrăştiere a scorurilor din jurul acestei tendinţe centrale.

Prin determinarea tendinţei centrale, se localizează *un singur punct* de pe scara de valori, care se materializează în medie, mediană sau modul. Atunci când determinăm dispersia, utilizând un indice oarecare de variabilitate, de fapt indicăm *o distanţă* de-a lungul scării de valori, *scoruri* sau *cote brute*.

Printre cei mai utilizaţi indicatori ai gradului de dispersie sau ai *variabilităţii*, sunt: ***rangul, abaterea medie, abaterea standard,*** şi ***variabilitatea.***

5.1. Rangul.

Cel mai simplu indicator al dispersiei este ***rangul*** sau ***amplitudinea.*** Acest indicator este determinabil prin simpla diferenţă dintre valorile extreme ale colecţiei de date brute.

$$R = X_{max} - X_{min} \qquad (5.1)$$

Revenind la exemplul prezentat la începutul capitolului de faţă, deci la cele două grupe de subiecţi, vom avea:

Grupul A: R=48-22 =26

Grupul B: R=58-20 =38

Concluzia care se desprinde imediat este că, în grupul B dispersia, împrăştierea valorilor, a cotelor este mai mare decât la grupul A.

Notăm faptul că, rangul este puţin utilizat, dar şi atunci cu prudenţă, deoarece este un indicator foarte variabil. Poate fi totuşi utilizat în cazul

grupelor mici. De asemenea, rangul poate induce erori grave în cazul în care între o valoare extremă şi majoritatea datelor există o distanţă prea mare, şi asta deoarece ignoră scorurile intermediare.

5.2. Abaterea medie

Un alt indicator al dispersiei este ***abaterea medie*** care reprezintă *media aritmetică a tuturor abaterilor de la medie a valorilor, ignorând semnele algebrice.*

Formula de calcul este:

$$AM = \frac{\sum |x - m|}{N} \qquad (5.2)$$

unde AM - abaterea medie

$|x - m|$ - abaterea de la medie fără a lua în considerare semnele algebrice.

Să luăm un **exemplu**: La un grup de 11 copii preşcolari s-a aplicat un test vizând vocabularul activ. Rezultatele obţinute sunt date în Tabelul 5.1.

Tabelul 5.1 Performanţele obţinute de cei 11 copii preşcolari la testul de vocabular.

Copii	Cota(x)	Diferenţă $\lvert x-m \rvert$	
Elena	20	20-14	6
Marta	18	18-14	4
Gheorghe	15	15-14	1
Vlad	15	15-14	1
Andrei	14	14-14	0
Mihai	14	14-14	0
Relu	14	14-14	0
Dimi	13	13-14	-1
Angela	13	13-14	-1
Alex	10	10-14	-4
Dan	8	8-14	-6
N=11	$\sum x=154$	$\sum \lvert x-m \rvert = 24$	

Înlocuind în Formula 5.2. datele noastre din Tabelul 5.1., vom obţine:

$$AM = 24/11 = 2,18$$

Acest indicator al dispersiei sau al variaţiei este mai rar utilizat în cercetările psihopedagogice sau sociologice datorită operaţiilor greoaie şi introducerii unei aproximări, mai ales când este vorba de un număr mare de date.

5.3. Abaterea standard

Cea mai mare popularitate în domeniul dispersiei o are ***abaterea standard*** care se notează σ (sigma) sau DS, dar în literatura de limbă engleză frecvent se notează S.

Se preferă acest indicator deoarece este exprimat în aceleaşi unităţi ca şi cele pe care le prelucrăm.

Notă: abaterea standard se calculează numai în cazul distribuțiilor de frecvente normale.

Precizăm că în cazul calculării dispersiei, a împrăștierii valorilor din jurul mediei se face distincție între:

A. ***Varianță*** (σ^2,DS^2, S^2) care este suma deviațiilor variabilelor de la medie la pătrat și împărțită la numărul de cazuri, simbolizat prin:

$$\sigma^2 = \frac{\sum (x-m)^2}{N} \qquad (5.3)$$

B. ***Abaterea standard*** (σ, ***DS, S***) care nu este altceva decât rădăcina pătrată din varianță, cu alte cuvinte din media pătratului abaterilor de la media măsurătorilor. Formula este:

$$\sigma = \sqrt{\sigma^2}$$

Înlocuind vom avea:

$$\sigma = \sqrt{\frac{\sum (x-m)^2}{N}} \qquad (5.4)$$

C. În literatura de specialitate se face distincție între: N și N-1 de la numitorul formulelor 5.3 și 5.4. Diferența este că: N reprezintă o **populație**, iar N-1 caracterizează un **eșantion** extras dintr-o populație mult mai mare. În același timp, N-1 este practic numărul de valori, de abateri de la medie care sunt dispuse să varieze (liber). Acest număr de valori libere care variază este numit ***grad de libertate***, despre care vom discuta într-un alt capitol.

În concluzie, formulele 5.3 și 5.4 pot fi exprimate și astfel în cazul eșantioanelor:

$$\sigma^2 = \frac{\sum (x-m)^2}{N-1} \qquad (5.5)$$

$$\sigma = \sqrt{\frac{\sum (x-m)^2}{N-1}} \qquad (5.6)$$

3.1. Abaterea standard la eșantioane mici

Atunci când avem un volum mic de date, prezentate aleator, calcularea abaterii standard se face utilizând Formula 5.6.

De exemplu: La măsurarea timpului de reacție al unui subiect s-au obținut datele prezentate în Tabelul 5.2.

Tabelul 5.2. Performanțele obținute de un subiect la determinarea timpului de reacție

Timpi de reacție x	(x-m)	$(x-m)^2$
20	20-18,8=1,2	1,44
17	17-18,8=-1,8	3,24
15	15-18,8=-3,8	14,44
18	18-18,8=-0,8	0,64
18	18-18,8=-0,8	0,64
22	22-18,8=3,2	10,24
16	16-18,8=-2,8	7,84
24	24-18,8=5,2	27,04
18	18-18,8=-0,8	0,64
19	19-18,8=0,2	0,04
24	24-18,8=5,2	27,04
21	21-18,8=2,2	4,84
15	15-18,8=-3,8	14,44
14	14-18,8=-4,8	23,04
18	18-18,8=-0,8	0,64
17	17-18,8=-1,8	3,24
20	20-18,8=1,2	1,44
23	23-18,8=4,2	17,64
N=18		$\sum(x-m)^2$=158,52

$$\sigma = \sqrt{\frac{158,52}{17}} = \sqrt{9,32} = 3,05$$

$$\sigma = 3,05$$

Concluzia: abaterea standard a timpului de reacţie a subiectului examinat este:

$\sigma = 3,05$

Menţionăm că, abaterea standard se calculează cu două zecimale.

Exemplul 5.1. Ne propunem să calculăm abaterea standard pentru datele din Tabelul 5.2.

1. Creăm variabila TIMP cu valorile din tabel (coloana x).
2. Din M **Statistics**, submeniul **Summarize**, O **Descriptives...** se completează câmpul **Variables** cu variabila TIMP
3. Se apasă butonul **Options** şi se bifează **Std. Deviation** din zona **Dispersion** pentru calculul abaterii standard
4. Se salvează fişierul cu numele timp.sav

Observaţie: Calculul abaterii standard se poate face şi cu ajutorul M **Statistics**, submeniul **Summarize,** O **Frecquencies**. Tot ca mai înainte, se trece variabila TIMP în lista de variabile, se apasă butonul **Statistics...**, iar aici se va bifa **Std. Deviation** din zona **Dispersion** pentru calculul abaterii.

5.3.2. Abaterea standard la eşantioane mari

Când volumul de date este mai mare, iar anumite valori se repetă de mai multe ori, este nevoie să utilizăm *gruparea* frecvenţelor. În aceste situaţii formula de calcul a abaterii standard se va modifica puţin la numărător în forma:

$$\sum (x-m)^2 = \sum fu \cdot u - \frac{(\sum fu)^2}{N} \qquad (5.7)$$

formula devine:

$$\sigma = \sqrt{\frac{\sum fu \cdot u - \frac{(\sum fu)^2}{N}}{N-1}}. \qquad (5.8)$$

În cele ce urmează vom exemplifica modul de utilizare a formulei 8, plecând de la distribuţia de frecvenţe prezentată în Tabelul 5.3

După cum se poate lesne constata în Tabelul 5.3, pe lângă coloanele noscute a mai apărut o coloană, a 5-cea. Ca şi definiţie de construcţie, oloana a cincea este produsul dintre coloana a treia (u) cu a patra (fu).

abelul 5.3. Performanţele obţinute de 83 subiecţi la un test de relaţii aţiale.

Variabila x	Frecvenţa f	Codarea u	fu	$fu\,u$
12	3	4	12	48
11	4	3	12	36
10	10	2	20	40
9	14	1	14	14
8	18	0	58 00	00
7	14	-1	-68 -14	14
6	10	-2	-20	40
5	6	-3	-18	54
4	4	-4	-16	64
i=1	Σf=N=83		Σfu=-10	Σfuu=310

Nu mai rămâne decât să facem înlocuirile în formula de calcul cu atele Tabelului 5.3.

$$\sigma = \sqrt{\frac{310 - \frac{(-10)^2}{83}}{82}} = \sqrt{3,76} = 1,94$$

Deci: $\sigma = 1,94$

5.3.3. Abaterea standard prin valori codate

În majoritatea situațiilor, suntem nevoiți să utilizăm pe lângă distribuția d frecvență și o grupare în intervale de clase a valorilor, utilizând un oarecare. În această nouă situație formula de calcul a dispersiei es modificată, în sensul că apare în plus și valoarea intervalului de clasă, at pentru varianță cât și pentru variabilitate.

Astfel, formulele în cauză devin:

$$\sigma^2 = \frac{i^2(\sum fu \cdot u - \frac{(\sum fu)^2}{N})}{N-1} \qquad (5.9)$$

și

$$\sigma = \sqrt{\frac{i^2(\sum fu \cdot u - \frac{(\sum fu)^2}{N})}{N-1}} \qquad (5.10)$$

Exemplu: În compoziția unei baterii de probe psihologice, pentr conducătorii auto, a fost și o probă de timp de reacție. Ca și performanță s avut în vedere exactitatea reacțiilor. În Tabelul 5.4 sunt date aces performanțe la 110 conducători auto examinați.

Tabelul 5.4. Performanțele la timp de reacție, exactitate la cei 1 conducători auto.

Variabila x	Frecvența f	Codarea u	fu	$fu\,u$
1,00-0,94	2	4	8	32
0,93-0,87	3	3	9	27
0,86-0,80	8	2	16	32
0,79-0,73	18	1	18	18
			51	
0,72-0,66	39	0	00	00
			-82	
0,65-0,59	16	-1	-16	16
0,58-0,52	13	-2	-26	52
0,51-0,45	6	-3	-18	54
0,44-0,38	3	-4	-12	48
0,37-0,31	2	-5	-10	50
i=0.07	Σf=N=110		Σfu=-31	Σfuu=329

După inspectarea vizuală a Tabelului 5.4 și a formulei de calcul a aterii standard 5.10 se va constata că se poate trece la înlocuire și deci m obține:

$$\sigma = \sqrt{\frac{0,07^2(329 - \frac{(-31)^2}{110})}{109}} = \sqrt{0,014} = 0,11$$

Rezultatul final este: σ: 0,11

5.4. Corecția lui Sheppard

Practica zilnică de cercetare , cu precădere în sociologie, dar și în hologie sau pedagogie ne pune în fața unei colecții foarte mari de date e, care sunt grupate pe baza unui interval de grupare mare. În aceste uații se introduce și o eroare în calcule numit "eroare de grupare", care buie descifrat și anihilat. În astfel de cazuri sau când dorim să lucrăm cu o mare exactitate se utilizează o formulă de corecție, practic de calculare a abaterii standard, "corecția lui Sheppard". Formula de calculare este :

$$\sigma_c = \sqrt{\sigma^2 - \frac{i^2}{12}} \qquad (5.11)$$

Să luăm în considerare un exemplu, unde: σ= 10,11 și un interval de up: i=21. Dacă efectuăm recalcularea cu formula lui Sheppard vom obține:

$$\sigma_c = \sqrt{102,21 - \frac{21^2}{12}} = 8,08$$

Deci, în cazul în care s-a lucrat pe eșantionul examinat s-a obținut o = 10,11, la un i =21. Intervalul de clasă fiind mare, efectul erorii de pare și-a făcut simțită prezența. Recalculând cu formula lui Sheppard s-a it valoarea obiectivă a coeficientului de variabilitate care este: σ_c = 8.09 - abatere standard corectată).

5.3.5. Combinarea dispersiilor şi a abaterilor standard

Aşa cum s-a arătat în subcapitolul referitor la media aritmetică există posibilitatea combinării a două sau mai multe medii provenite de la grupe diferite. Tot aşa, există posibilitatea aglutinării dispersiilor sau abaterilor standard provenite de la diferite grupe plecând de la datele individuale ale grupelor. Pentru acest lucru se utilizează formula:

$$\sigma_a^2 = \frac{(n_1-1)\sigma_1^2 + (n_2-1)\sigma_2^2 + n_1 d_1^2 + n_2 d_2^2}{N-1} \qquad (5.12)$$

unde $n_1 + n_2 = N$ numărul de subiecţi

$M_1 - M = d_1$ diferenţa dintre media grupului 1 şi media generală

$M_2 - M = d_2$ diferenţa dintre media grupului 2 şi media generală

Menţionăm faptul că, această formulă se poate extinde la câte grupe dorim să combinăm.

5.4. Interpretarea abaterii standard

În primul rând trebuie să facem distincţie între : abaterea standard a unei populaţii notată σ si abaterea standard obţinută într-un eşantion *s* care a fost extras dintr-o populaţie oarecare.

De asemenea, variabilitatea se poate referi în principal la două situaţii distincte:

- ***variabilitatea interindividuală***, atunci când este vorba de diferiţi subiecţi din eşantion, exemplul din Tabelul 5.3;
- ***variabilitatea intraindividuală***, când avem o serie de date ale aceluiaşi individ, exemplul din Tabelul 5.2.

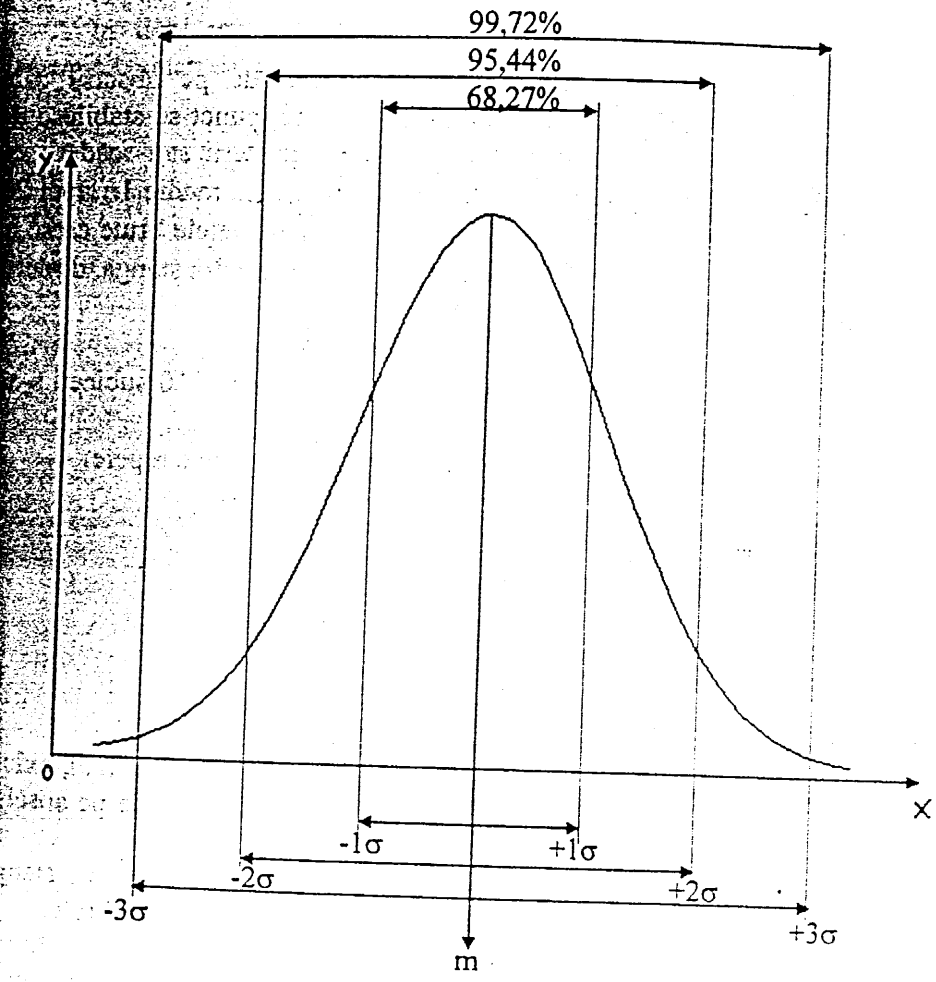


Fig. 5.2. Prezentarea grafică a interpretării abaterii standard

Abaterea standard poate fi utilizată şi ca unitate de măsură pe abscisa unei distribuţii de frecvenţă. În acest caz interpretarea abaterii standard se face în termeni procentuali. În Figura 5.2. este vizualizată această interpretare.

Utilizarea abaterii standard ca şi unitate de măsură se bazează pe faptul că se exprimă în aceleaşi unităţi ca şi cotele brute pe care le avem în

calcule. Locul de la care se pleacă este punctul de pe abscisă c corespunde mediei eşantionului examinat. De la acest punct se stabilesc regulă cele şase puncte de reper corespunzătoare creşterii sau scăderii câte o abatere standard (Fig. 5.2). Astfel, adăugându-se la medie 1σ apoi şi in final 3σ se obţin punctele din dreapta mediei unde datele brute cresc valoare absolută. Scăzând 1σ, 2σ şi 3σ se obţin punctele din stânga, direc descreşterii cotelor brute ca şi valoare absolută.

Exemplu: Într-un experiment psihologic unde s-au folosit 110 subiecţi s obţinut următoarele date: m=60,2 şi σ=3,50.

Plecând de la cele menţionate vom obţine următoarele repere:

-3σ = 47,70
-2σ = 53,20
-1σ = 56,70
m = 60,2
+1σ = 63,70
+2σ = 67,20
+3σ = 70,70

Cu alte cuvinte între medie şi plus o abatere standard (63,70), exis o distanţă, o întindere de 3,50 unităţi brute, iar punctul 67,20 de pe absc se află la şapte unităţi brute de medie.

În acelaşi mod se procedează şi pentru reperele aflate în stân mediei. Astfel, punctul 47,70 se află la o distanţă de 10,50 unităţi brute.

Deci, abaterea standard în sensul arătat mai sus, măsoară distanţe care se află o cotă oarecare în raport cu media din cadrul unui eşanti examinat.

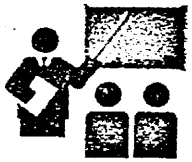
Al doilea aspect legat de aceste puncte de reper este interpreta procentuală. S-a demonstrat faptul că, între aceste repere există un procen constant. În felul acesta, între +1σ şi -1σ se află aproximativ două trei din totalul cazurilor, mai precis 68,27% (Fig. 5.2.). Între +2σ şi -2σ se a un procent de 95,44% din totalul cotelor brute, iar între +3σ şi -3σ situează 99,72% din totalul observaţiilor aflate în analiză.

Revenind la exemplul nostru, vom spune că între +1σ şi -1σ , ad între 56,70 şi 63,70 se află 68,27% din subiecţi (au obţinut cote brute în aceste limite) adică un număr de 75 subiecţi din totalul de 110.

Identic se procedează şi cu intervalele +2σ şi -2σ şi +3σ şi -3σ.

În final menţionăm, încă o dată, faptul că abaterea standard se determină, calculează numai în cazul distribuţiilor normale, simetrice.

Capitolul 6 CORELAȚIA



- **coeficientul de corelație Bravais Pearson**
- **coeficientul de corelație**
- **diagrama de corelație**

6.1. *Noțiunea de corelație*

În capitolele anterioare s-au prezentat unele metode de sistematiza
și prelucrare primară a datelor aparținând unei **singure variabile.**
practică, foarte des, suntem puși în situația de a determina **posibila legătu**
existentă între două sau mai multe variabile. Această posibilă relație
stabilește prin ***calculul corelației.*** După Mărgineanu (1938), "corelația pu
unitate și ordine în ceea ce pare izolat și dispersat". Prin studiul corelați
putem stabili relația dintre: înălțime și greutate la un grup de subiec
legătura dintre performanțele la un test de inteligență și randamentul șco
etc.

Studiul relației dintre variabile are ca punct de plecare *variat*
simultană sau ***covariația*** acestora. Prin corelație căutăm să determină
modul și gradul de legătură între variabilele în cauză.

De la început trebuie subliniat faptul că: *studiul corelației*
dezvăluie legătura de cauzalitate ci, așa cum s-a arătat, este modul
covariere, asociere a variabilelor în cauză.

De asemenea, în studiile experimentale relațiile sunt unidirecționa
(x determină y). În cazul corelațiilor, variabilele sunt date, nu manipulate,
relația nu este vectorizată.

<u>De retinut</u>: în calculul corelației fiecare subiect din eșantion trebu
să posede **cel puțin două valori** sau măsurătorile luate în studiu să se refe
la **perechi de subiecți.**

Practica statistică a determinat dezvoltarea mai multor tipuri
corelații, specifice situațiilor concrete. Alegerea tipului potrivit de corela
se face ținând cont de următorii factori :

- tipul de scală de măsurare utilizată pentru fiecare variabilă (sc
nominală, ordinală, de intervale sau de proporții);
- natura distribuției în cauză(continuă sau discretă);

caracteristicile distribuției de frecvență;

Calculul corelației pleacă de la tabelul de corespondență, care
ține subiecții (liniile tabelului) și variabilele(coloanele x, y etc.) sau
ușirile aflate în studiu (Tabelul 6.1.)

2. Diagrama de corelație

Așa cum un set de date brute poate fi vizualizat grafic, tot așa și
ația dintre două variabile poate fi vizualizată. Imaginea intuitivă a
elației dintre două variabile este ***diagrama de corelație*** (Figura 6.1.).

emplu: Unui lot de 15 elevi i s-au administrat două probe: un test de
morie și un test de cunoștințe de matematică. Se pune întrebarea dacă
performanțele obținute la cele două teste există sau nu o legătură.
belul 6.1 redă grafic relația existentă între cele două variabile.

belul 6.1. Performanțele obținute de cei 15 elevi la cele două teste.

Subiecți	Test de memorare variabila x	Test de cunoștințe variabila y
ENA	48	8
OGDAN	50	10
IHAI	44	6
IRCEA	43	5
EORICA	47	5
UCIA	49	7
INA	44	7
ARMEN	48	9
EX	45	6
OREL	47	7
NGELA	48	10
HEORGHE	44	3
ARIA	45	5
OICU	42	3
RMAND	46	6

După cum se poate remarca din Tabelul 6.1, fiecărui subiect îi
respund câte două valori: o cotă brută reprezentând testul de memorare
ariabila x) și câte o

În mod obişnuit, pe axa 0x sunt trecute valorile brute reprezent
variabila independentă în timp ce pe axa 0y sunt cotele variab
dependente.

În mod obişnuit, pe axa 0x sunt trecute valorile brute reprezent
variabila independentă, în timp ce pe axa 0y sunt cotele variab
dependente.

Pentru reprezentarea grafică a relaţiei dintre cele două variabile
folosim de cele două axe rectangulare 0x şi 0y. Marcăm pe cele două
valorile corespunzătoare cotelor brute, respectiv cea a notelor, având grij
plasăm valorile spre punctul 0 (Figura 6.1).

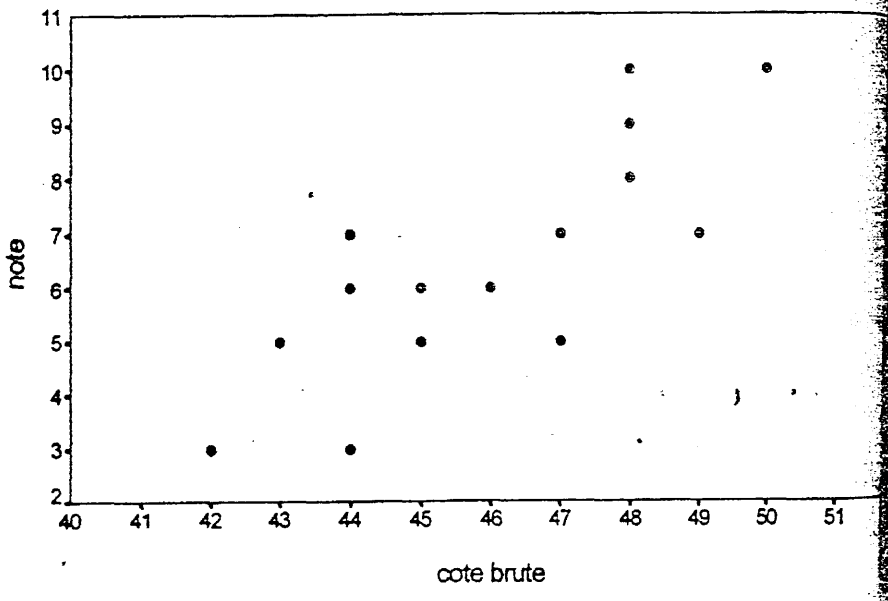


Figura 6.1. Diagrama de corelaţie dintre testul de memorare (x) şi ce
cunoştinţe (y).

Exemplul 6.1. Să trasăm diagrama de corelaţie pentru datele din Tabelul 6.1 relativ la corelaţiile dintre rezultatele obţinute la testul de memorare şi testul de cunoştinţe.

1. Se crează un nou fişier cu variabila COTA pentru rezultatele testului de memorare şi NOTA pentru rezultatele testului de cunoştinţe.
2. Se completează cele două variabile cu valorile corespunzătoare din tabel.
3. Se trasează diagrama din M **Graph** O **Scatter...**, se alege tipul **Simple.**
4. Se apasă butonul **Define**; în câmpul **Y Axis** se trece variabila NOTA, iar în câmpul **X Axis** se trece variabila COTA.
5. Se confirmă acţiunea cu **OK.**
6. Dacă dorim să modificăm titlurile abscisei, respectiv ale ordonatei (**cota** → cote brute şi **nota** → note), în fereastra de **Chart** care conţine diagrama, se face un dublu click cu mouse-ul pe textul **cota.**
7. În fereastra nou apărută **X Scale Axis** se modifică: câmpul **Axis Title** cu noul text *Cote brute*; în zona **Major Division**, câmpul **Increment** se modifică din 2 în valoarea 1, iar în zona **Minor Division**, deasemenea câmpul **Increment** se modifică din 2 în valoarea 1.
8. Identic, pentru a modifica **nota** → note, se face un dublu click cu mouse-ul pe textul **nota.**
9. În fereastra nou apărută **Y Scale Axis** se modifică: câmpul **Axis Title** cu noul text *Note*, în zona **Major Division**, câmpul **Increment** se modifică din 2 în valoarea 1, iar în zona **Minor Division**, deasemenea câmpul **Increment** se modifică din 2 în valoarea 1.
10. Optional, se poate salva diagrama din M **File** O **Save As...** cu numele corel.cht.
11. Se salvează fişierul de date cu numele corel.sav.

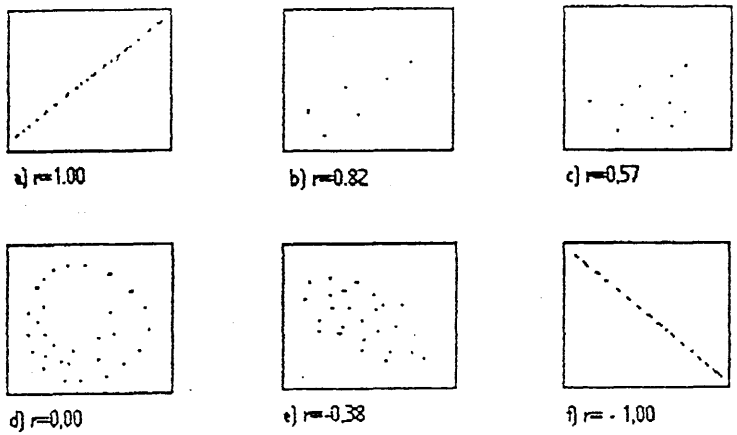


Fig. 6.2. Diferite grade de legătură dintre variabile (Elifson, 1990)

Corelaţia pozitivă se obţine în situaţiile în care cele două setur date variază în acelaşi sens (Fig. 6.2, a,b,c). Situaţia prezentată în Figura este un exemplu tipic de corelaţie pozitivă, adică performanţele cres ambele variabile; cota ridicată la variabila x atrage după sine o cotă ma nivelul variabilei y.

Corelaţia negativă, Figura 6.2, e,f, este caracteristică situaţiilo care cele două variabile au tendinţe inverse de dezvoltare. Cu alte cuvin aceste relaţii în timp ce o variabilă creşte, cealaltă variabilă tinde să sc De exemplu: relaţia dintre numărul de greşeli şi exactitate, cu cât num de greşeli este mai mare cu atât mai mic va fi coeficientul de exactitate.

Corelaţia independentă sau zero apare atunci când între cele variabile nu există nici o legătură liniară (Figura 6.2., d).

Sintetizând cele arătate până acum, putem afirma: *coeficientu* ***corelaţie*** *este un indice care exprimă legătura liniară dintre două vari Acest indice poate varia între: +1,00 şi -1,00 trecând prin valoarea* Figura 6.2

6.3 Calculul coeficientului de corelaţie

Statistica din ştiinţele sociale cunoaşte mai multe tipuri de corelaţii, dintre acestea cele mai uzuale sunt: *coeficientul de corelaţie Bravais - Pearson* (notat *"r"); coeficientul de corelaţie al rangurilor* (notat ρ, care se iteşte: ro); *coeficientul de corelaţie biserial şi triserial; coeficientul de corelaţie multiplu* etc.

În capitolul de faţă ne vom ocupa doar de primele două: Bravais - Pearson şi al rangurilor. Celelalte tipuri de corelaţii şi altele, vor fi prezentate pe larg în volumul următor al cursului de faţă.

6.3.1. Coeficientul de corelaţie Bravais - Pearson (r).

Cel mai frecvent utilizat coeficient de corelaţie în psihologie, sociologie, pedagogie şi nu numai este coeficientul Bravais - Pearson notat cu *r.*

În cele ce urmează vom detalia două metode de calculare ale acestui coeficient de corelaţie.

6.3.1.1 Metoda abaterii medii.

Este vorba de o metodă mai rar utilizată în practică datorită volumului mare de muncă şi a efortului pe care-l comportă. Menţionăm că este o metodă recomandată când volumul de date este mic şi se utilizează calculatorul.

Formula de calcul a coeficientului de corelaţie, prin utilizarea metodei abaterii medii este:

$$r = \frac{\sum (x-m)(y-m)}{\sqrt{[\sum (x-m)^2][(y-m)^2]}} \qquad (6.1)$$

Σ ??

Exemplificarea utilizării acestei formule o vom face în baza date prezentate în Tabelul 6.1. Folosind aceste date compunem Tabelul 6.2.

Tabelul 6.2. Procedura de calcul al corelației prin utilizarea meto abaterii medii.

Variabila			Variabila			
x	x-m	$(x-m)^2$	y	y-m	$(y-m)^2$	(x-m)(y-m)
48	2	4	8	1,6	2,56	3,2
50	4	16	10	3,6	12,96	14,4
44	-2	4	6	-0,4	0,16	0,8
43	-3	9	5	-1,4	1,96	4,2
47	1	1	5	-6,4	40,96	-6,4
49	3	9	7	0,6	0,36	1,8
44	-2	4	7	0,6	0,36	-1,2
48	2	4	9	2,6	6,76	5,2
45	-1	1	6	-0,4	0,16	0,4
47	1	1	7	0,6	0,36	0,6
48	2	4	10	3,6	12,96	7,2
44	-2	4	3	-3,4	11,56	6,8
45	-1	1	5	-6,4	40,96	6,4
42	-4	16	3	-3,4	11,56	13,6
46	0	0	6	-0,4	0,16	0
m=46,0			m=6,4		$\Sigma(y-m)^2=65,8$	$\Sigma(x-m)(y-m)$ =57

După cum se poate observa, Tabelul 6.2 ne pune la dispoziție to datele de care avem nevoie pentru a putea calcula valoarea coeficientului corelație utilizând Formula 6.1.

$$r = \frac{57}{\sqrt{78 \cdot 65,8}} = 0,796$$

Deci, valoarea coeficientului de corelație este: r = 0,796, c obișnuit se scrie: r = .796. Momentan ne mulțumim să afirmăm că, ac

coeficient de corelație este foarte semnificativ, dar interpretarea coeficientului de corelație va fi prezentată într-un subcapitol viitor.

Exemplul 6.2 Calculul coeficientului de corelație Bravais-Person pentru datele din Tabelul 6.2.

1. Se crează variabilele VARX și VARY cu datele din tabel.
2. Din M **Statistics** O **Correlate** Submeniul **Bivariate** se aleg variabilele VARX și VARY în câmpul **Variables.**
3. Pentru calculul coeficientului de corelație Bravais Pearson se bifează opțiunea **Pearson** din zona **Correlation Coefficients.**
4. Opțional, matricea de corelație obținută se poate salva din M **File**, O **Save SPSS Output** cu numele r1.lst.
5. Se salvează fișierul de date cu numele de r1.sav.

Se va obține:

```
          VARX        VARY

VARX       1,0000       ,7960
          (   15)     (   15)
           P= ,        P= ,000

VARY        ,7960      1,0000
          (   15)     (   15)
           P= ,000     P= ,

(Coefficient / (Cases) / 2-tailed Significance)
```

6.3.1.2. Metoda produselor.

Cea mai folosită metodă de calcul al corelației este cea a produselor. Formula de bază este:

$$r = \frac{N(\sum xy) - (\sum x)(\sum y)}{\sqrt{[N(\sum x^2) - (\sum x)^2][N(\sum y^2) - (\sum y)^2]}} \qquad (6.2.)$$

În continuare vom exemplifica modul de utilizare al Formulei (6.2.) Pentru acest lucru vom compune Tabelul 6.3.

x	y	x^2	y^2	xy
8	11	64	121	88
6	15	36	225	90
9	19	81	361	171
12	23	144	529	276
5	10	25	100	50
23	33	529	1089	759
16	15	256	225	240
10	14	100	196	140
14	18	196	324	252
9	21	81	441	189
21	29	441	841	609
Σx=133	Σy=208	Σx^2=1953	Σy^2=4452	Σxy=2864

Tabelul 6.3. Calculul coeficientului de corelație utilizând formula de bază

În continuare nu rămâne decât să înlocuim în Formula 6.2 date concrete din Tabelul 6.3. Astfel, vom obține:

$$r = \frac{(11 \cdot 2864) - (133 \cdot 208)}{\sqrt{[(11 \cdot 1953) - (133)^2][11 \cdot 4452) - (208)^2]}} = 0{,}825$$

După cum se vede, valoarea coeficientului de corelație este: r =.82

6.3.1.3. Metoda valorilor codate

În practica cotidiană cercetătorii din domeniul psihologiei sociologiei sau al pedagogiei sunt confruntați cu situații în care numărul cote, scoruri este foarte mare. După cum se știe, aceste situații per ***gruparea*** datelor în clase în baza unui interval de grupare și astfel se po

ela la o procedură simplificată de calcul *a coeficientului de corelație r, zat pe operația de codare* pe care am mai întâlnit-o .

Exemplu: Un cercetător este interesat să afle dacă între capacitatea decizie și cunoștințele tehnice la conducătorii de autovehicule ofesioniști există vreo legătură. În consecință a aplicat un chestionar ând probleme tehnice și un chestionar prin care să clasifice atitudinea cizională. Cunoștințele tehnice au fost notate cu cote între 4 și 12, în timp atitudinea decizională cu valori între -13 (imprudență) și 21 (atitudine cesiv de prudentă).

În Tabelul 6.4 sunt trecute performanțele celor 67 de conducători o examinați.

belul 6.4. Performanțele celor 67 de conducători auto profesioniști la estionarul tehnic (variabila y) și atitudinea decizională (variabila x).

y	x	y	x	y	x	y	x	y	x	y	x	y
12	-3	9	-8	9	-1	10	222	9	-6	10	-10	10
9	1	7	2	12	-10	8	-3	11	2	7	-2	8
7	2	8	8	6	6	7	6	8	10	9	1	7
6	11	6	1	8	14	9	13	8	-1	6	9	9
9	-9	10	-5	7	0	5	15	7	19	8	10	7
8	20	5	4	9	7	10	-10	5	4	10	-6	8
7	8	6	10	5	21	6	9	8	20	6	16	10
8	2	8	-13	8	2	8	7	7	16	7	N=67	
6	12	4	11	4	3	5	6	8	18	8		
11	-4	9	0	11	-12	9	-2	10	-7	9		

După cum se poate vedea din Tabelul 6.4 fiecare subiect are câte două cote, corespunzătoare performanțelor la cele două chestionare. Plecând la aceste date se întocmește o distribuție de frecvență în care să fie prezentate ambele variabile ale subiecților. Astfel, se întocmește Tabelul cu două intrări, pe orizontală apare distribuția variabilei x, iar pe verticală apare desfășurarea variabilei y. Ușor se poate remarca faptul că

acest tabel este o simplă diagramă de corelaţie (în interiorul cheiarului elementele distribuţiei de frecvenţe pe orizontală pentru variabi (atitudinea decizională), iar pe verticală pentru variabila y (cunoş tehnice).

Tabelul 6.5. Determinarea lui r utilizând tehnica grupării şi a codării da brute.

Variab	*Variabila y*									f_x	u	
x	4	5	6	7	8	9	10	11	12	6	3	
21-17		1	2		3					7	2	
16-12	1			2	2	1	1			12	1	
11-7	1	1	3	3	1	2	1			15	0	
6-2		1	1	2	5	3	1	1	1	13	-1	
1- -3		1	2	2	2	2	2	2		7	-2	
-4 - -8			2	1	3	1				7	-3	
-9 - -13		1			2	1	2		2		-3	
f_y	2	5	8	11	16	12	8	3	2	N=67		Σf
u	-4	-3	-2	-1	0	1	2	3	4			
$f_y u$	-8	-15	-16	-11	00	12	16	9	8	$\Sigma f_y u$=	-21	
$f_y uu$	32	45	32	11	00	12	32	27	32	$\Sigma f_y uu$ =	22	
$nu_x u_y$	-12	0	-14	-1	0	-1	-14	-6	-12	$\Sigma nu_x \cdot u_y$=	-6	
	-8	-9	-12	-4		2	4	0	0			
	-4	-3	-6	-3		2	2	-6	-12			
		0	0	0		0	0					
		3	4	2		-2	-4					
		9		4		-6	-4					
						3	-12					

Singura noutate a Tabelului 6.5 este linia: $\Sigma nu_x \cdot u_y$ (este vorb suma algebrică produsului liniilor u ale celor două variabile trecute tabelul de diagramă). Cu alte cuvinte, această linie se obţine din s algebrică a produsului dintre valorile codate ale celor două variabil frecvenţa corespunzătoare din diagrama de corelaţie. De exemplu: în coloanei corespunzătoare cotei 4 a variabilei y se obţine:

- (-4) • (1) • (2) = -8
- (-4) • (1) • (1) = -4

- Σ = (-8) plus (-4) = -12, valoare ce se înscrie la locul potrivit.

Identic se procedează şi cu celelalte coloane, iar în final se face suma brică a acestor valori.

Utilizând datele obţinute în Tabelul 6.5 se poate calcula valoarea ficientului de corelaţie cu ajutorul formulei:

$$r = \frac{\sum nu_x \cdot u_y - \frac{(\sum f_x u)(\sum f_y u)}{N}}{\sqrt{(\sum f_x u \cdot u - \frac{(\sum f_x u)^2}{N})(\sum f_y u \cdot u - \frac{(\sum f_y u)^2}{N}}} \quad (6.3.)$$

ocuind în Formula 6.3. datele concrete din Tabelul 6.5 se va obţine:

$$r = \frac{-60 - \frac{(-4)(-21)}{67}}{\sqrt{(198 - \frac{(-4)^2}{67})(223 - \frac{(-21)^2}{67})}} = -0{,}284$$

ci, valoarea coeficientului de corelaţie obţinut este: r = -0,284.

6.2. Coeficientul de corelaţie al rangurilor.

În unele situaţii, respectiv în urma efectuării unor investigaţii, zultatele obţinute sunt sub formă de clasificare prin rang. Aceste situaţii r de regulă, la utilizarea scalelor ordinale de măsură. În aceste situaţii formanţele subiecţilor nu pot fi cuantificate, dar se pot face o ierarhizare funcţie de un criteriu precis definit.

Verificarea existenţei covarierii între două variabile aflate în astfel situaţii este posibilă prin utilizarea unei tehnici speciale, respectiv ilizând formula lui ρ (ro) pentru determinarea coeficientului de corelaţie.

Aceeaşi metodologie este recomandată şi în cazul când efectivul antionului este mai mic decât 30. În aceste situaţii cotele brute sunt locuite cu ranguri.

Exemplu: Într-un experiment, H. Suhrweier şi D. Sonderschule (196 studiat relaţia dintre acuitatea vizuală şi rezultatele la învăţătură. în copiilor ambliopi. În Tabelul 6.6 sunt prezentate rezultatele finale

Tabelul 6.6. Procedeul de calculare al coeficientului de corelaţie ρ.

Subiecţi	Rangurile		d	d^2
	variabila x	variabila y		
A	1	7	6	36
B	2	1	1	1
C	3	3	0	0
D	4,5	5,5	1	1
E	4,5	5,5	1	1
F	6	12	6	36
G	7,5	10	2,5	6,25
H	7,5	8	0,5	0,25
I	9	10	1	1
K	10	10	0	0
L	11	2	-9	81
M	12	4	-8	64
N=12				Σd^2=227,5

Calculul coeficientului de corelaţie se face cu ajutorul formulei:

$$\rho = 1 - \frac{6\sum d^2}{N(N^2 - 1)} \quad .....(6.4)$$

De la prima inspectare a Tabelului 6.6 ne putem da seama de fa că, ea cuprinde toate elementele pentru calculul corelaţiei prin interme Formulei 6.4

227,5

$12(144-1)$

În final s-a obţinut coeficientul de corelaţie: ρ=0,210.

Exemplul 6.3. Vom exemplifica calculul coeficientului de corelaţie al rangurilor pentru datele din Tabelul 6.6 folosind coeficientul de corelaţie Spearman.

1. Se crează variabilele VARX şi VARY.
2. Din **Statistics** O **Correlate** Submeniul **Bivariate** se aleg variabilele VARX şi VARY în câmpul **Variables.**
3. Pentru calculul coeficientului de corelaţie al rangurilor se bifează opţiunea **Spearman** din zona **Correlation Coefficients.**
4. Optional, se poate salva matricea de corelaţie cu numele de r2.lst.

va obţine:

```
- - - SPEARMAN CORRELATION COEFFI
IENTS - - -

ARX                ,1947
                N(   12)
                Sig ,272

                VARY

Coefficient / (Cases) / 1-tailed Significance)
```

6.4. *Interpretarea coeficientului de corelaţie.*

Aşa cum s-a arătat la începutul capitolului, prin corelaţie pu
stabili:

- dacă există sau nu o relaţie de covariaţie între două variabile;
- direcţia, sensul acestei legături (pozitivă sau negativă);
- coeficientul de corelaţie va indica şi gradul de asociere a celor d
variabile (maximă ±1 sau lipsă 0);

J.I.Guilford a stabilit cinci limite de semnificaţie a valor
coeficienţilor de corelaţie:

1. valorile între: .00 şi ± .20 aproape neglijabile;
2. valorile între: ± .21 şi ± .40 prezentă, dar mică;
3. valorile între: ± .41 şi ± .70 moderată, semnificativă;
4. valorile între: ± .71 şi ± .90 corelaţie ridicată;
5. valorile între: ± .91 şi 1,00 foarte semnificative.

Pentru a evita unele confuzii ce apar în legătură cu interpreta
coeficientului de corelaţie sunt necesare câteva precizări:

- Coeficienţii de corelaţie *r* şi *ρ* se interpretează identic, ambele referin
se la covariaţie.
- Orice corelaţie diferită de zero este semnificativă, stabileşte un **anu**
grad de relaţie între cele două variabile aflate în studiu.
- Gradul de relaţie **nu este** proporţional cu mărimea coeficientului
corelaţie. Exemplu: coeficientul de corelaţie *r* =.60 **nu** indică o relaţie
două ori mai mare faţă de *r* =.30.
- Orice corelaţie trebuie privită întotdeauna în relaţie cu situaţia în ca
fost obţinută. Cu alte cuvinte, mărimea coeficientului de corelaţie
reprezintă o valoare dată, absolută. De aici necesitatea specificării
mai exacte şi clare a variabilelor aflate în studiu. Orice coeficient
corelaţie este raportat la circumstanţele în care a fost obţinut şi v
interpretat doar în lumina acestor circumstanţe.

Coeficientul de corelaţie evidenţiaza daca între doua variabile există sau
nu o legătură, respectiv covariaţie. Corelaţia este o legătură statistică şi
nu una de natură cauzală.

Practic coeficienţii de corelaţie se stabilesc plecând de la eşantionare.
Dar ne interesează, de fapt, în ce fel coeficientul de corelaţie *r* sau *ρ*
caracterizează colectivitatea din care s-a extras eşantionul respectiv.
Pentru acest lucru ne folosim de **coeficientul de corelaţie critic.** Acest
coeficient reprezintă **valoarea minimă** ce trebuie să o ia un coeficient de
corelaţie obţinut pentru a fi semnificativ.

Pentru a uşura interpretarea coeficientului de corelaţie din acest
ct de vedere s-a întocmit un tabel special (Anexa 1) care conţine
orile critice pentru diferite praguri de semnificaţie şi pentru diferite valori
ii *n* (*n* = *N* - 2).

Menţionăm că tabelul conţine valorile **absolute** ale coeficientului
ic. Astfel, tabelul poate fi utilizat şi în cazul când coeficientul de
elaţie este negativ.

emplu: Utilizând metoda abaterii medii în cadrul acestui capitol s-a
inut: *r* = .796, N = 15.

În primul rând stabilim valoarea lui n. n = N - 2 deci n = 15 - 2 = 13.
Tabelul valorilor critice vom găsi:

n	p=.10	p=.05	p=.02	p=.01
13	.44	.51	.59	.64

Valoarea coeficientului găsit este: r =.796 (obişnuit se scrie .796)
te este mai mare decât cea critică la gradul de probabilitate p = .01.
ncluzia: corelaţia, gradul de legătură dintre cele două variabile aflate în
diu este foarte mare şi semnificativă. Acest lucru se scrie condensat:
r = .796, *p* < .01.

Revenind la nivelul experimentului vom putea concluziona: între
pacitatea de memorare şi nivelul cunoştinţelor din matematică există o
rte strânsă relaţie.

Reamintim, este vorba de o **relaţie** statistică, cauzalitatea rela
trebuie căutată la nivelul analizei relaţiilor dintre variabilele independen
dependente din cadrul investigaţiei **experimentale.**

La exemplificarea utilizării Formulei 6.3 s-a obţinut: r = - 0,28
un eşantion N = 67.

Paşii algoritmului de interpretare vor fi:

- Stabilirea lui n: n = N-2; adică: n = 67 -2 =65;
- Identificarea valorilor critice din tabel:

n	p=.10	p=.05	p=.02	p=.01
...				
70	.20	.23	.27	.30

Valoarea 65 nefiind în tabel ne vom referi la valoarea lui *n* cea m
apropiată, în cazul nostru 70.

- Concluzia: coeficientul de corelaţie obţinut este semnificativ
  un prag de probabilitate: p < .02
- Concluzie generală: între capacitatea de decizie şi cunoştinţe
  tehnice există o strânsă relaţie **negativă** sau **inversă** *r* = -.28
  p < .02. Adică, creşterea uneia din variabile duce la scăder
  celeilalte variabile.
- Analiza cauzalităţii se face la nivelul variabilelor din cadr
  experimentului derulat.

Atunci când s-a prezentat tehnica de calcul a coeficientului
corelaţie al rangurilor s-a obţinut o valoare: ρ =.210 şi N = 12. Interpretar
coeficientului de corelaţie ρ parcurge aceleaşi paşi ca şi în caz
coeficientului Bravais-Pearson, adică:

- Stabilirea lui n, care de această dată este: *n* = N-2; în caz
  nostru:
  n = 12-2=10;
- Identificarea în tabelul valorilor critice:

n	p=.10	p=.05	p=.02	p=.01
10	.50	.58	.66	.71

Concluzie: conform valorilor critice, valoarea coeficientului calculat, ρ
=.210 este cu mult sub aceste limite. Deci, valoarea este nesemnificativă.
Concluzie generală: între acuitatea vizuală şi rezultatele la învăţătură nu
există nici o legătură (lipseşte factorul comun) în cazul copiilor
ambliopi.

În final, subliniem că aceşti coeficienţi de corelaţie au o deosebită
importanţă în stabilirea validităţilor şi a fidelităţilor. De asemenea, utilitatea
coeficienţilor de corelaţie a fost demonstrată în toate domeniile de
cercetare ale psihologiei.

ANDRIEŞ C. LUCIAN
Nr 2774
Biblioteca Personală

Exercițiul 1.

Momentul central în conducerea unui autovehicol îl constituie capacitatea de decizie, materializată comportamental printr-o atitudine ofensivă sau defensivă. Practic, este vorba de o singură trăsătură, dar care este bipolară, având la un pol atitudinea ofensivă, riscantă (cote brute negative), iar la celălalt pol situându-se atitudinea defensivă , mergând până la o prudență exagerată. (valori pozitive ale cotelor).

Între cele două extreme se situează conducătorii auto echilibrați din punct de vedere al atitudinii decizionale (în general aceștia acoperă aproximativ 66% din eșantionul examinat).

Exploatând cunoștințele dobândite până în prezent, determinați pentru eșantionul prezentat limitele între care se situează cele trei tipuri de atitudini decizionale: ofensivi, echilibrați, defensivi.

14,25	2,67	5,00	-13,00	-1,00	1,33
9,43	43,00	5,33	31,33	-0,57	-10,00
48,40	11,00	14,00	10,80	-11,00	17,50
3,00	-17,00	3,00	-4,40	-8,67	-1,00
8,67	-29,00	16,50	8,29	3,00	4,74
4,00	-7,33	18,50	-4,20	-5,00	-1,33
-2,67	0,00	-2,67	-12,50	-13,60	9,50
60,50	-2,00	-5,50	13,00	-7,00	-3,50
59,00	9,25	0,00	-3,33	-0,25	-3,50
-11,00	18,50	12,67	9,00	21.33	2,50

18,50	-13,50	9,00	0,00
-11,50	5,50	-3,33	3,00
12,67	2,67	13,00	-2,00
9,00	14,25	-12,50	2,00
9,50	-5,33	-4,20	9,50
21,33	0,30	8,29	-0,25
2,50	16,50	-4,40	-1,33
-29,00	-13,60	10,80	1,00
-4,40	17,50	31,33	1,00
0,00	0,00	-0,57	-17,00

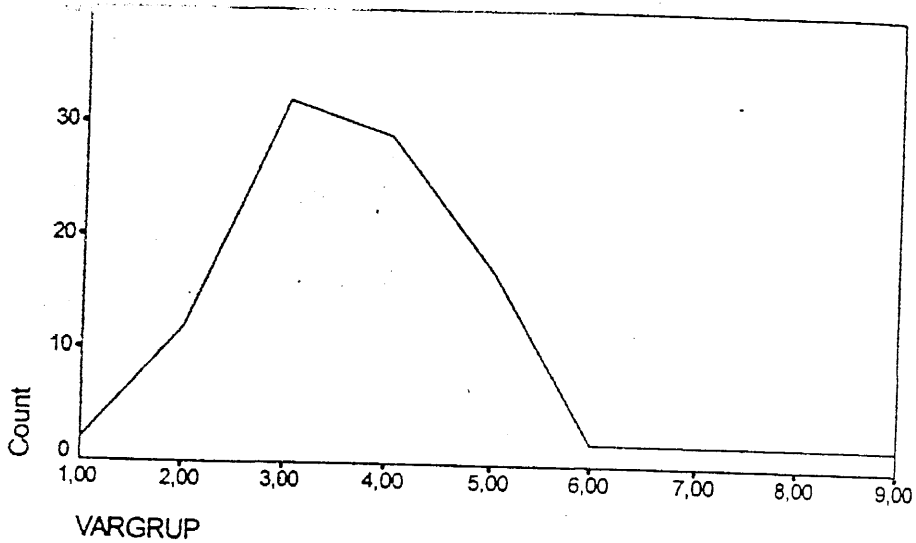
N=100.

1. Se crează variabila VAR
2. Din M **Statistics** submeniul **Summarize** se alege O **Frecquencies**. Se introduce variabila VAR în câmpul **Variables**.
3. Se apasă butonul **Statistics...**; pentru calculul indicilor tendinței centrale (medie, mediană, mod) se vor bifa opțiunile **Mean, Median** și **Mode** din zona **Central Tendency**, iar pentru indicii dispersiei se vor bifa **Std. Deviation, Variance** și **Range** din zona **Dispersion.**
4. Se salvează fișierul cu numele auto.sav.

Se va obține:

Mean	4,142	Median	1,665	Mode	,000
Std dev	14,765	Variance	217,992	Kurtosis	3,800
S E Kurt	,478	Skewness	1,326	S E Skew	,241
Range	89,500				

1. Se trasează poligonul frecvențelor (folosind ...)



2. Se fac comentarii privind forma curbei obținute, prin utilizarea noțiunilor de: Kurtosis, Skewness și Range.
3. Utilizând cunoștințele de la interpretarea abaterii standard, se calculează apoi se comentează relațiile: $m \pm \sigma$ (obținându-se cele trei tipuri comportamentale: riscant, echilibrat, prudent).
4. Totul se trece într-un raport de cercetare care se listează.

adecvată situației în cauză.
Consemnați concluziile comparative într-un raport de cercetare.

Variabila	Frecvențe			
	Bărbați		Femei	
	Alcoolici	Nealcoolici	Alcoolici	Nealcoolici
66-71	1			
60-65	6		3	
54-59	13	1	2	1
48-53	13	1	10	2
42-47	17	3	11	1
36-41	33	3	12	1
30-35	32	2	8	8
24-29	32	9	11	17
18-23	23	16	5	26
12-17	24	36	2	40
6-11	7	43	2	49
0-5	1	25		21

1. Pentru comparație, utilizați metoda grafică.
2. Comentați tendințele centrale.

Exercițiul 3.

În cadrul unei baterii de probe psihologice, au fost folosite următoarele probe: I test de vocabular, II test matematic, III spirit de observație, IV test de relații spațiale și un criteriu general V.

Rezultatele sunt date în tabelul A.

Se cere:

1. Comentarea compoziției bateriei de probe psihologice și a relației cu criteriul utilizat.
2. Condensarea concluziilor într-un raport de cercetare.

Tabelul A. Rezultatele obținute la probele bateriei psihologice și la criteriu.

Test I	Test II	Test III	Test IV	Criteriu
22	11	24	39	30
8	5	22	23	21
19	6	44	12	21
32	8	72	20	33
13	2	25	20	29
24	5	30	6	8
22	4	38	37	35
35	1	54	28	16
18	7	37	46	30
13	10	61	46	35
53	23	56	41	38
15	9	42	5	18
34	18	30	58	46
15	2	42	21	17
27	4	28	26	24
19	9	32	13	19
29	4	24	0	27
24	9	42	21	23
27	9	54	20	30
16	5	42	24	34

Test I	Test II	Test III	Test IV	Criteriu
[illegible]	12	[illegible]	40	26
22	5	58	41	20
32	4	57	4	16
18	8	49	36	42
24	15	87	34	26
22	12	14	16	27
22	10	38	0	20
21	21	32	43	37
13	10	52	35	11
23	3	60	13	37
2	10	29	21	27
20	4	50	8	27
25	11	76	20	26
14	6	40	8	46
11	2	32	4	26
2	9	61	10	20
38	17	56	20	35
16	6	61	23	21
14	4	17	7	21
23	25	61	29	16

ANDRIEȘ C. LUCIAN
Nr 2774
Biblioteca Personală